Title
Essays in Labor and Demographic Economics

Author
Yen, Maysen

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Labor and Demographic Economics

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Maysen Yen

Dissertation Committee:
Professor David Neumark Irvine, Chair
Professor Matthew Freedman
Professor John Duffy

2021

DEDICATION

To

my wife Ling Ling, and to my parents Shyh-Jaung and Jion for all their support and encouragement

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank, first and foremost, my committee chair David Neumark. I am very thankful for all the guidance and advice he has given to help me develop into an economist and researcher. Additionally, I am very grateful for all the professional opportunities David has given me, such as Graduate Student Researcher funding, collaborating on papers together (including the 2nd chapter of this dissertation), and providing me with the opportunity to get involved with economic consulting where I could apply economic analyses to real-world problems.

I would also like to thank Matthew Freedman and John Duffy for their advice on various parts of the dissertation chapters. Matthew helped me tremendously with my numerous questions on accessing and using Census data while John was able to provide me with valuable feedback and support towards developing my first real research project and experiment.

The text of Chapter 2 in this dissertation is a reprint of the material as it appears in *Demography*. The co-author (David Neumark) listed in this publication directed and supervised research which forms the basis for that chapter in this dissertation. Springer Nature has also granted permission for use of that publication in this dissertation.

Finally, I would like to thank the financial support given to me by the Department of Economics and the School of Social Sciences.

CURRICULUM VITAE

Maysen Yen

Education

University of California, Irvine
PhD in Economics, 2021

University of California, Irvine
MA in Economics, 2016

University of Wisconsin – Madison
BA, Economics and Political Science, 2013

Professional Experience

Economic Consultant
Self-Employed, 2018 – 2021

Business Analyst
Axos Bank, 2015

Financial Analyst
Kimberly-Clark Corporation, 2013 – 2014

Research Experience

Graduate Research Assistant
University of California, Irvine, 2017 – 2021

Undergraduate Research Assistant
University of Wisconsin – Madison, 2011 – 2013

Teaching Experience

Teaching Assistant
University of California, Irvine, 2015 – 2017, 2020

Publications

Neumark, David and Maysen Yen. 2020. "Relative Sizes of Age Cohorts and Labor Force Participation of Older Workers," Demography 57(1): 1-31.

ABSTRACT OF THE DISSERTATION

Essays in Labor and Demographic Economics

By

Maysen Yen

Doctor of Philosophy in Economics

University of California, Irvine, 2021

Professor David Neumark, Chair

My dissertation's primary contribution is identifying factors that affect differences in labor market outcomes among various demographic groups.

In the first chapter, I designed an experiment to examine the effects of stereotype threat, a concept from psychology where people conform or feel at risk of conforming to negative stereotypes about their group, under competition. I found that providing information about performance differences as a stereotype threat changed the subjects' preferences to favor the tasks that their gender is perceived to perform relatively better in but did not affect performance outcomes. The implication is that even if stereotype threat does not affect one's ability to perform, it may affect education and career choices that could contribute to group differences in the labor market.

The second chapter was written jointly with Dr. David Neumark and published in *Demography*. This paper analyzed the effects of cohort sizes on the labor force participation and wages of older workers in the United States. Older workers increased their labor force participation as their relative size to the working age population increases, which is contrary to the standard labor supply hypothesis. Additionally, when using a richer model that accounts for the size of older workers relative to younger workers, we found the demand for older workers was high when their cohort size is large relative to

prime age workers, suggesting that older workers enter in more flexible working arrangements in their later age.

Finally, the last chapter analyzed the effects of expanding public transit infrastructure on labor market outcomes in Los Angeles. I use panel data on tracts, treating route placement as endogenous, which is then instrumented by the distance from the centroid of each tract in LA to a hypothetical Metro route. Overall, I find proximity to Metro stations increases labor force participation and employment for residents, which is robust to using both a binary and continuous measure of distance. Additionally, I find evidence that increased job density in neighborhoods near new transit stations is contributing to the employment increase.

1. Does Stereotype Threat Affect Beliefs and Preferences More than Outcomes?

1.1 Introduction

There are persistent gender differences in labor market outcomes despite progress towards equality in the workplace. Common reported statistics include the fact that 5% of CEOs in S&P 500 companies are women (CNN) or that women make 83% of men's weekly wages (Bureau of Labor Statistics). Economists have historically attributed gender differences in labor market outcomes to discrimination, (Becker, 1957; Hellerstein et al., 1999; Black and Strahan, 2001), occupational self-selection (Polachek, 1981), or behavioral differences.

A growing literature in experimental labor economics explores behavioral differences between men and women, particularly regarding gender differences in individuals' willingness to compete and performance outcomes under competition. Gneezy et al. (2003) found that men perform better in a competitive tournament payoff structure while Niederle and Vesterlund (2007) found that when given a choice, women tend to shy away from competition by choosing a non-competitive piece-rate payoff structure over the tournament.[1] Some researchers have argued that task choice matters, and there are inherent gender perceptions in math or logical puzzles that could lead to lower outcomes in competition for women. Thus, recent work also used tasks that are perceived to be female-dominated and do not find evidence that women underperform in tournaments for these type of tasks (Shurchkov, 2012; Iriberri and Rey-Biel, 2017). Combined, this evidence suggests that perceptions of gender gaps in the tasks rather than a lower overall willingness to compete play a more significant role in explaining gender differences in observed outcomes.

The differences in results across the two types of tasks suggests that stereotype threat may be the culprit for lower outcomes under competition for women in tasks where men are perceived to perform better. Stereotype threat is a concept originating in the social psychology literature where people are at risk, or feel that they are at risk, of conforming to a stereotype about their group (Steele and Aronson, 1995). If stereotype threat is present in the labor market, it may lead to differences in productivity through lower performance for the stereotyped as low-performing group, which could translate to lower wages or fewer promotions. This effect may be further amplified in competitive

[1] An extensive literature replicates and confirms the original Niederle-Vesterlund design (Healy and Pate, 2011; Kamas and Preston, 2012; Niederle et al., 2013).

environments, where only the top-performers are rewarded in the form of promotions or wage increases, disincentivizing effort for the negatively stereotyped group. Steel and Aronson (1995) performed the first experiment that invoked stereotype threat among Black students by referencing the GRE Verbal task as diagnostic of ability, which they claimed invoked negative stereotypes about Black intelligence, and found negative effects on performance. Spencer et al. (1999) mentioned to subjects that there were gender differences in mathematical performance as a stereotype threat and found that this generated a larger gender gap.

Following these two seminal papers, there have been numerous papers in social psychology and a few in economics exploring stereotype threats with race (Steele, 1997; Aronson et al., 1999; Stone et al., 1999), gender (Cadinu et al., 2005; Fryer et al., 2008), and socioeconomic status (Hoff and Pandey, 2006; Desert et al., 2009). In these studies, the stereotype threat was implemented by stating there were differences between groups or indirectly by mentioning intellectual ability (see Appendix Table B1 for examples of stereotype threats used in the literature). The stereotype threat phenomenon is not immune to criticism and skepticism. First, it may be subject to publication bias within the psychology literature, replication issues, and experimenter demand effects (Ganley et al., 2013; Fryer et al., 2008). [2]

This paper has four aims. First, this study provides a new, direct test of the effect of stereotype threat on economic performance by providing subjects with information on the actual performance of other subjects, drawn from the same subject pool, on the same performance task. Specifically, I provide information on the actual gender gap in scores that I observed from pre-treatment rounds involving sixty-four subjects to another set of subjects who were drawn from the identical population (University of California – Irvine undergraduate students) and I provide exact magnitudes of the score differences. Previous studies exploring the effect of stereotype threat have relied on implicit beliefs or vague statements about group differences to invoke the stereotype threat (see Appendix Table B1).[3] [4] Additionally, I show that subjects respond to the information about actual performance gaps by

[2] Fryer et al. (2008) pays subjects, which psychology experiments to not do, to minimize experimenter demand effects and do not find evidence of significant gender differences in performance from stereotype threat.

[3] For instance, a conventional approach to inducing stereotype threat is to mention that there are differences in mathematical and/or logical abilities between men and women (Spencer et al., 1999; Cadinu et al., 2005; Fryer et al., 2008).

[4] There are a few existing studies on how men and women differ in responses to feedback, but these are focused at the individual level (Berlin and Dargnies, 2016; Wozniak et al., 2014; Iriberri and Rey-Biel, 2017). Iriberri and Rey-Biel (2017) also frame their paper as a study on stereotype threat, but they are only using individual-level feedback. This is problematic since stereotype threat is traditionally understood as the performance of a group (Steele and Aronson 1995).

updating their beliefs about gender gaps, which makes this form of stereotype threat clear and convincing.

Secondly, this study explicitly considers how outcomes can be affected when magnitudes of subjects' perceptions (beliefs) about the gender gap are different from the magnitude of actual gender gaps. It is expected that subjects' beliefs will be further amplified or attenuated when reinforced or contradicted by information about group performances. There are two types of plausible scenarios for stereotype threat that specifically target females, which will be the focus of this study. For example, if a female believes there are small or no gender differences but is given information that there is a larger gender difference favoring males, then she may exert less effort from the stereotype threat and decrease her performance. Additionally, if a female believes that males outperform females on a specific task but finds out there is no gender gap, her performance may also suffer.[5] In this experiment, I ask about the magnitude of subjects' perceptions of gender gaps in the tasks, after the experiment but also during the experiment as a treatment to test these perceptions ex-ante and firmly ground these implicit beliefs by requiring subjects to reflect upon them.[6]

Thirdly, this study tests the effect of stereotype threat within a lab-based competitive environment. If a rational actor believes that there are group differences and that they will underperform relative to another group, they may exert less effort in tournaments relative to a piece-rate payment scheme. There are numerous behavioral mechanisms offered in the psychology literature as an explanation to the underlying mechanisms behind stereotype threat. Some explanations related to effort are that the stereotyped group may experience increases in self-handicapping strategies and increased dejection. Additionally, subjects may suffer from greater anxiety, reduced self-efficacy, and increased negative thinking and thought suppression (Pennington et al., 2016). While these reasons have not been compared between a competitive and non-competitive environment in the psychology literature, my prior is that the negative motivational explanations for stereotype threat listed above are stronger under a competitive tournament. This ties nicely into existing literature on gender gaps in

[5] Another scenario is if a female believes that males outperform females on a specific task but finds out that the gender gap is not as large as her prior belief, she may be more compelled to exert more effort and increase her performance. Stereotype boost is defined as when results are boosted from positive stereotypes about one's group and stereotype lift is defined as when results are boosted from negative stereotypes about another's group (Shih et al., 2011). In this study, I focus specifically on the two forms of stereotype threat for women, but other tasks could be chosen to test for stereotype boost.

[6] In the few existing studies that consider how the task can reinforce perceptions (Shurchkov, 2012; Iriberri and Rey-Biel, 2017), the researchers measure such perceptions by asking subjects whether they believe males or females performed better in the task in absolute terms after the experiment, and do not consider adjustments in the magnitudes of these differences.

competition, especially in the tasks where a credible stereotype threat can be elicited due to gender differences in performance.

Finally, a natural follow-up question would be whether these perceptions of gender gaps in performance, real or imagined, and subsequent information about actual gender differences among groups may subtly influence an individual's beliefs about one's own performance and task preferences. While stereotype threat is concerned about the effects on actual performances, little work has been done on how information about one's group performance affects one's preference for a specific type of task or one's confidence in their performance. For example, it is possible that if the information provision about gender gaps is credible, subjects may prefer the task their group is deemed better in or avoid tasks that their group underperforms on average. This could have important implications in the real world – for example, if women experience stereotype threat in STEM courses, they may be dissuaded from pursuing STEM careers, even if their performance in those courses does not suffer.

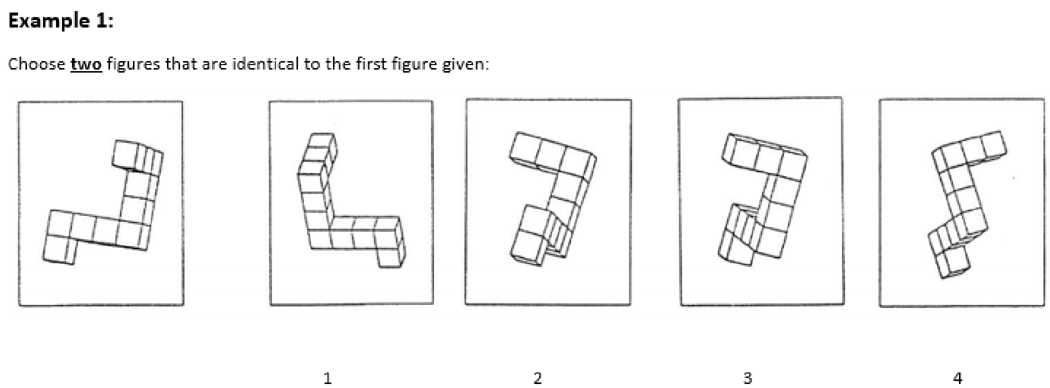
To summarize the results, I find that providing information on gender gaps shifted the subjects' preferences towards the task their group is reported to perform better in, suggesting behavioral changes are affected rather than performance outcomes. I do not find convincing evidence that the stereotype threat treatments played a role in affecting the performance of individuals in either the competitive or non-competitive setting. The rest of the paper is organized in the following manner. Section 1.2 covers the experimental design. Section 1.3 covers the results. Section 1.4 briefly concludes and discusses future avenues to explore.

1.2 Experimental Design

1.2.1 Overview of Design

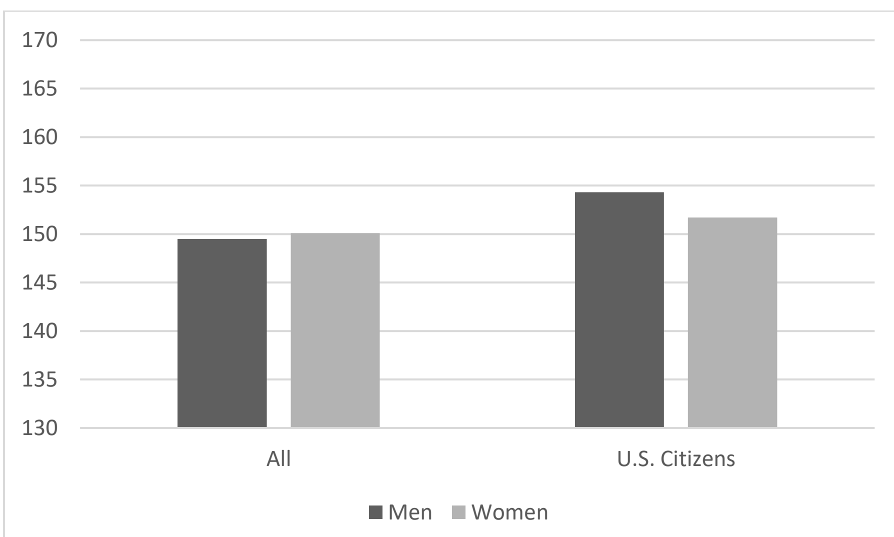
Prior to running the experiment, I conducted a pilot study and identified two credible tasks where I could impose a credible stereotype threat. The first task was the mental rotation puzzle, because it was a task where men outperformed women while participants did not believe a gender gap exists. The mental rotation task relies on spatial skills, which has historically favored men (Iriberri and Rey-Biel, 2017; Shepard and Metzler, 1971; Maccoby and Jacklin, 1974). In this task, participants are asked to find two figures that are identical in shape to the first figure given (see Figure 1.1). The second task I identified as suitable was GRE Verbal questions, because it was a task where women were believed to outperform men, but the actual gender difference was negligible, serving as a good foil to the mental rotation puzzles.[7] For example, Educational Testing Service (ETS) reports show that scores are only slightly higher for women overall (Figure 1.2).[8] I modified these tasks to be suitable for use with the computer software, z-tree.

Figure 1.1 Mental Rotation Task



Participants are shown the following graphics and asked to enter in the corresponding numbers to two figures that are identical to the first figure. The answer to this one would be 1 and 3.

Figure 1.2. ETS Average GRE Verbal Scores by Gender (2016)



Source: Educational Testing Service.

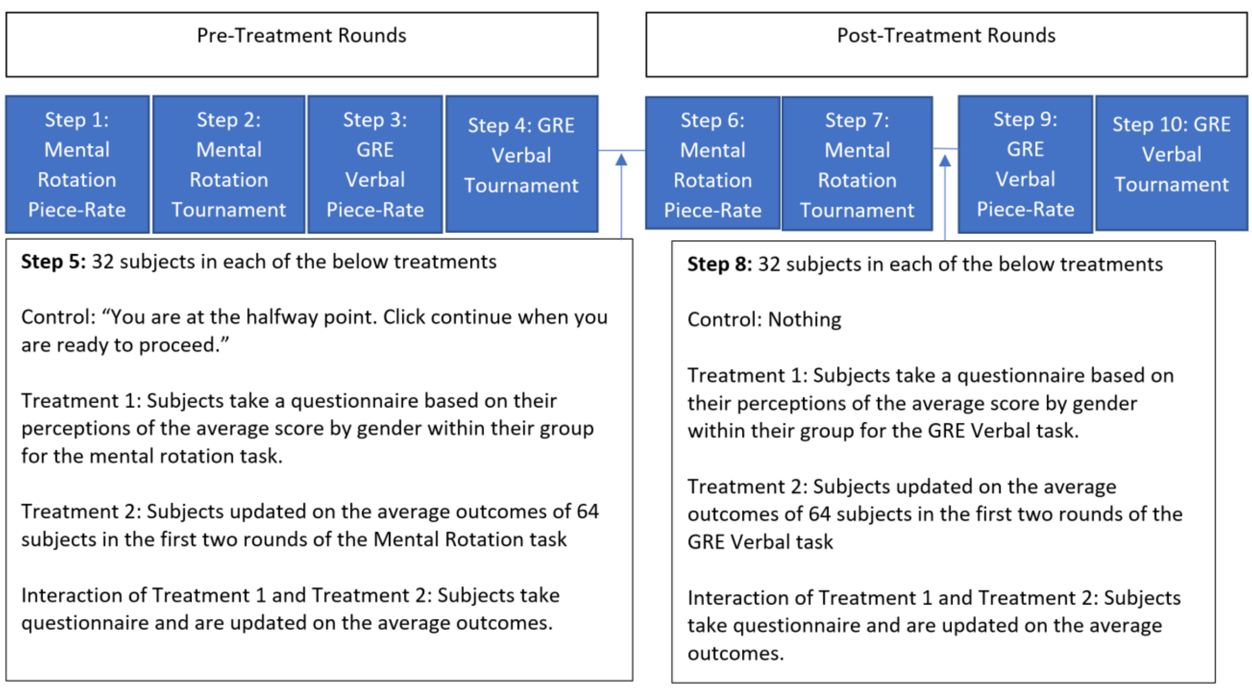
---

[7] GRE Verbal was chosen instead of the SAT Reading test used in the pilot study due to the ease of computerized implementation.

[8] Among U.S. citizens, which make up the bulk of my subjects (or at least to my knowledge, non-international students), men score higher. However, this is not a concern as the main condition for this task to contradict existing beliefs is that women should not score higher than men.

The experiment consisted of a total of 128 subjects, recruited through University of California – Irvine's Experimental Social Science Laboratory, conducted in sessions throughout November 2017 to January 2018. These subjects were broken down into groups of two men and two women. There were 32 subjects in each control group and treatment group. These groups were divided into eight groups of four. The gender composition of the four-person group was not announced explicitly but can be observed by the participants as they were divided into their groups and sat next to their assigned group numbers. Along with standard instructions, subjects were informed that they will be competing with subjects within their group in certain stages of the experiment. Each subject was given a $7 show-up payment. Each mental rotation round had 20 questions with 20 seconds for each question. The GRE Verbal task consisted of 8 questions in each round with either 30 or 45 seconds for each question depending on the question type. The total time per session was 90 minutes, broken into eight rounds of tasks, and the average total pay per participant was $14.89.[9]

Figure 1.3. Experimental Design



Order of whether piece-rate or tournament goes first will be randomized. They will always be in sequential order. For example, half of the subjects will follow the order listed above while the other half will start with a tournament and then piece-rate subsequently.

[9] For more details on the experiment instructions, see Appendix C for screenshots

Figure 1.3 demonstrates the sequence of events and different treatments. The experiment consisted of eight rounds, four pre-treatment rounds, which were identical for all groups, and four-post treatment rounds, which varied depending on the treatment variables. In the pre-treatment phase, two rounds were the mental rotation task (Step 1 & 2) and two rounds were the GRE Verbal task (Step 3 & 4). These rounds were performed under either a piece-rate scheme or a tournament, assessing competition in a within subjects design. The order was randomized with half the group starting with the piece-rate and half the group starting with the tournament to account for learning within each task. Under the piece-rate scheme, subjects had the ability to earn an additional $0.10 for each mental rotation question they answer correctly and $0.25 for each GRE Verbal question they answered correctly. Under the tournament scheme, they won an additional $4 in that round if they are the highest scoring in their group, which was split in the event of a tie. For all treatment groups, the first four rounds (Steps 1-4) are identical.

The experiment is set up as a 2x2 design. In Step 5, I imposed the treatments for the mental rotation task first. For the control group, 32 subjects are simply told they are at the halfway point. Treatment 1, the perception treatment, asked 32 subjects to answer questions on what they believe the average pre-treatment scores of the men and women in their group is for the mental rotation task. They were incentivized to answer honestly by having the person who guessed the closest to the actual values to earn an additional $4. Some of the previous studies relied on implicit beliefs about group differences to generate a stereotype threat, such as mentioning ability or invoking a rival's gender. The purpose of the perception treatment was to mimic previous studies in that it is reliant of implicit beliefs but also to ask the magnitude of the perceived difference. Asking about magnitudes is important, because it 1) gives us a measure of how large these implicit beliefs about gender gaps are, 2) allows comparisons with actual performance outcomes, and 3) has the potential to invoke a stereotype threat when these perceived gaps are large. This question addresses whether reminding students of their own biases' and perceptions will elicit changes in their performance in subsequent rounds. Since subjects did not perceive a gender difference in the mental rotation task, the perception treatment should not affect the future rounds for either gender.

Treatment 2, the update treatment, provided 32 subjects information on the observed gender gap of a sample of 64 subjects' pre-treatment scores for the mental rotation task. For clarification, this logistically amounted to providing subjects the average pre-treatment scores of the 64 subjects in the control group and Treatment 1 group.[10] Information provision is stronger than treatment 1 in that it explicitly informed subjects what the average gender gap was in each of the tasks. This allowed us to measure the effects of how updating information on actual gender differences affected future outcomes. If stereotype threat exists as a behavioral phenomenon, then notifying subjects of a large gender gap in favor towards men in the mental rotation task should decrease the performance of women in the post-treatment rounds. This is a novel treatment that improves upon other stereotype threat elicitations which only vaguely allude to gender differences (Spencer et al.,1999; Cadinu et al., 2005; Fryer et al., 2008) but do not give additional details about the magnitude of the differences nor mention that the difference was generated by an identical task in an identical lab setting.

To my knowledge, the only stereotype threat study to provide subjects with such direct information on the task they were performing was by Iriberri and Rey-Biel (2017). In one of their treatments, subjects were updated on their own performance and/or a rival's performance and they interacted this treatment with a stereotype threat treatment in which subjects were reminded of their rival's gender. However, by having only having two participants in a tournament and revealing the outcome of the other participant, their study uses individual feedback rather than information about group characteristics, which is not a reflection of stereotype threat since it does not deal with average outcomes. Rather, the information about a rival's gender appeared to be the main trigger for stereotype threat in their study. Thus, the information provision in my study is an improvement, because it provides a realistic group average that is more consistent with our understanding of stereotype threat, especially when related to the language of early studies where they allude to group abilities (Aronson et al., 1999; Spencer et al., 1999; Cadinu et al., 2005; Fryer et al., 2008). Some may argue an individual may believe that they are "better than average" and overestimate their ability relative to their group (Moore and Healy, 2008), thereby lessening the effects of the information provision treatment. However, stereotypes have always been about group outcomes and this phenomenon should have been present in all stereotype threat studies.

Finally, I interacted the perception treatment and the update treatment. 32 subjects took the questionnaire first as described in Treatment 1, and then are immediately given updated information about the gender gap as described in Treatment 2. This primes and grounds subjects' perceptions of gaps directly and then immediately reinforces or contradicts these perceptions with score outcomes.

---

[10] All groups have identical pre-treatment settings. From a practical perspective, I only had information from the first 64 subjects sampled at the time of running treatment 2.

In the mental rotation task where perceptions on the gap will not be as high as the actual gap in performance, interacting these two treatments may intensify the stereotype threat and further decrease performance for women, particularly in the competitive rounds.

After subjects are either subjected to the settings in the control group, treatment 1, treatment 2, or both treatments, they are given more mental rotation task questions in both the piece-rate and tournament payoff style (steps 6 and steps 7). It is expected that the treatment variables affect subjects' performance in these rounds. Step 8 then imposed the same treatments for the GRE Verbal task. In treatment 1, I predict that since the perception of the gender gap is large in favor of women, men may decrease performance in the tournament round in the presence of stereotype threat. In treatment 2, by updating men that they did not underperform relative to women, I expect a boost in men's post-treatment scores. The interaction of these treatments may capture a positive, differential effect, but this will not be significant if perceptions are firmly rooted. Following these treatments, subjects are given additional GRE Verbal questions to solve (Step 9 and Step 10).

Following these steps, subjects took a post-experiment questionnaire that collected other demographic characteristics about the subjects, such as anxiety, race, age, task preferences, payment preference, willingness to compete, and which task they believed they performed better in. I also collected perception of the gender gap at this stage as well, but for all four tasks. Subjects who were exposed to Treatment 1 were already asked their perception of the gender gap for the pre-treatment rounds, and this perception is not expected to change by much. However, for subjects exposed to Treatment 2, this belief elicitation allows us to test whether subjects perceived the information given as credible. If subjects believed the information given, their perception of the gender difference at this stage should reflect the gender gap that was provided.

Subjects were then informed of their total payment only, which was the cumulative total of all eight rounds.[11] Subjects were not given feedback between rounds; the winner of the tournaments was not announced to the subjects nor were they given information about their relative rank. Subjects were then paid and dismissed.

---

[11] Charles et al. (2016) offers a detailed literature review of the benefits of paying for one round versus paying for all rounds. The authors acknowledge that "pay all" tends to be the baseline used in the literature and that there is no clear best method. While paying for only one round may reduce wealth effects, there is also the potential for diluted incentives and the introduction of background risk. Thus, for the purposes of this experiment, I paid all the subjects for every round.

1.2.2 Designing the information provision treatment from pre-treatment scores

A comparison of the pre-treatment scores will motivate our initial discussion. As stated in the experimental design and from my priors from the pilot study, men are expected to outperform women in the mental rotation task and men and women are expected to perform about equally in the GRE Verbal task. Table 1.1.A shows the average score of the pre-treatment rounds, where all 128 subjects were exposed to identical tasks under the same conditions. Males perform better than females in both the mental rotation task by 2.563 points (out of 20 points) and the GRE Verbal task by 0.313 points (out of 8 points). When comparing pre-treatment scores by either piece-rate or tournament payoff styles, the differences between men and women are similar in magnitude for the mental rotation task. The GRE Verbal task paints a slightly different story. Here, the gender difference is 0.516 in favor of males for the tournament style (p-value of 0.024 from a two-sided t-test) but the difference was smaller and not statistically significant for the piece-rate.

From a practical perspective, to construct treatment 2 with the updated information about score differences, I had to use a smaller pool of pre-treatment data from subjects in the control group and treatment 1 with the available information at the time. Thus, for the information provision in treatment 2, there was data from 64 subjects available, shown in Table 1.1.B. The results from the limited pool was slightly different than the full group. Here, the mental rotation difference was only 1.594 points on average in favor towards males (p-value of 0.081 from a two-sided t-test). The GRE Verbal was 0.125 points in favor towards males, although this is not statistically significant. If this experiment were to be repeated, I would use the scores from the full sample in Table 1.1.A to construct the information provision treatment. It is unclear if reporting the higher gaps in the full sample would generate a stronger effect than the limited pool. Nevertheless, it was the observed gender gap available at the specific time.[12]

[12] It is possible to incorporate additional pre-treatment scores of the subjects in Treatment 2 to a total of the average scores for 96 subjects in real-time. For example, among the 32 subjects in Treatment 2 and the 32 subjects receiving both Treatment 1 and 2, I could have updated the pre-treatment scores with the average scores of the current subject pool. However, I chose not to do this to keep the magnitudes consistent for both groups, as essentially, I would be varying the treatment among the groups receiving Treatment 2 and both Treatment 1 and Treatment 2 depending on the performance of the current subjects.

Table 1.1: Actual Performance during Pre-Treatment Rounds

	(1) Men	(2) Women	(3) Difference	(4) p-values	(5) # men	(6) # women
<u>**A. All Subjects**</u>						
Average of Both Rounds						
Mental Rotation	11.813	9.25	2.563***	(0.000)[0.000]	64	64
GRE Verbal	2.320	2.008	0.313*	(0.115)[0.082]	64	64
Piece Rate Round						
Mental Rotation	11.953	9.438	2.516***	(0.000)[0.000]	64	64
GRE Verbal	2.156	2.047	0.109	(0.662)[0.445]	64	64
Tournament Round						
Mental Rotation	11.672	9.063	2.609***	(0.000)[0.000]	64	64
GRE Verbal	2.484	1.969	0.516***	(0.024)[0.016]	64	64
<u>**B. First 64 Subjects**</u>						
Average of Both Rounds						
Mental Rotation	11.203	9.609	1.594*	(0.081)[0.091]	32	32
GRE Verbal	2.109	1.984	0.125	(0.674)[0.674]	32	32
Piece Rate Round						
Mental Rotation	11.344	9.688	1.656*	(0.092)[0.080]	32	32
GRE Verbal	2.063	2.031	0.032	(0.936)[0.690]	32	32
Tournament Round						
Mental Rotation	11.063	9.531	1.531	(0.111)[0.113]	32	32
GRE Verbal	2.156	1.934	0.219	(0.498)[0.508]	32	32

Mental Rotation Score is out of 20 and GRE Verbal Score is out of 8 in each round. The p-values in parenthesis are two-sided t tests while the brackets are the Mann-Whitney U test for non-parametric assumptions. Section B only contains the 64 subjects from the control group and perception treatment. These results were used to construct treatment 2's stereotype threat language, in which actual performance data was given to the subjects.

Table 1.2. Perceptions on Pre-Treatment Rounds Average Performance

	(1) Men	(2) Women	(3) Difference	(4) p-values	(5) # men	(6) # women
A. All Genders' Beliefs						
Mental Rotation	12.317	12.244	0.073	(0.775)[0.810]	32	32
GRE Verbal	4.739	5.186	-0.447***	(0.001)[0.001]	32	32
B. Men's Beliefs						
Mental Rotation	12.697	12.578	0.119	(0.746)[0.550]	32	0
GRE Verbal	4.625	5.194	-0.569***	(0.006)[0.007]	32	0
C. Women's Beliefs						
Mental Rotation	11.938	11.909	0.028	(0.939)[0.849]	0	32
GRE Verbal	4.853	5.178	-0.325*	(0.091)[0.067]	0	32

Mental Rotation Score is out of 20 and GRE Verbal Score is out of 8 in each round. The p-values in parenthesis are two-sided t tests and the brackets are the Wilcoxon signed-rank test for non-parametric assumptions. The 64 subjects are drawn from those who participated in the perception treatment, where they were asked about their perception of the average score by gender in their group.

1.3 Results

1.3.1 Do subjects believe the updated information?

I first check whether the updated information treatment was deemed credible by the subjects. If this information is credible, it should affect subjects' own beliefs about gender differences in the tasks in the post-experiment questionnaire where all subjects are asked about their perceived average scores by gender within their group (for all the rounds). In the update treatment, subjects were told the information in Table 1.1.B. (men performed 1.6 points better than women and that men and women performed about equally in the GRE Verbal task, although men performed 0.125 points better). If subjects believed these results, those exposed to the update treatment will trend towards these results. The regression is as follows:

$$Belief^{G,T} = \alpha + \delta_1^{G,T} Perception + \delta_2^{G,T} Update + \delta_3^{G,T} Perception * Update + \epsilon_{it} \qquad (1.3.1)$$

Belief captures the subject's perceived gap of the score difference of all rounds (in our regression context, it is the subjects' belief of the average men's scores minus their beliefs of the average women's scores). G is a superscript that denote gender where G ={M, F} for males and females, T = {R, V} for Mental Rotation or GRE Verbal. Perception and Update are the treatment variables described above that take on a value of 1 if treated, 0 otherwise.

Table 1.3: Post-Experiment Belief About Gender Gap

Group	(1) All	(2) Men	(3) Women
A. Mental Rotation			
Perception Treatment	-0.238	-0.05	-0.425
	[0.439]	[0.633]	[0.627]
Update Treatment	1.344***	1.500**	1.188*
	[0.439]	[0.633]	[0.627]
Perception Treatment*Update Treatment	-0.034	-0.075	0.006
	[0.621]	[0.895]	[0.886]
constant	0.219	0.063	0.375
	[0.310]	[0.447]	[0.443]
N	128	64	64
R-Squared	0.133	0.152	0.119
B. GRE Verbal			
Perception Treatment	0.297	0.650**	-0.056
	[0.244]	[0.263]	[0.400]
Update Treatment	0.625**	0.875***	0.375
	[0.244]	[0.263]	[0.400]
Perception Treatment*Update Treatment	-0.428	-0.706*	-0.15
	[0.345]	[0.372]	[0.565]
constant	-0.500***	-0.938***	-0.063
	[0.172]	[0.186]	[0.283]
N	128	64	64
R-Squared	0.057	0.19	0.023

The dependent variable is the difference between the self-reported beliefs about the average scores of men minus the self-reported beliefs about the average scores of women in the group.

Table 1.3 provides evidence that subjects understood and believed the updated information. The interpretation of the results from Table 1.3.A is that exposure to the updated information on average, increased the belief that men outperformed women by 1.344 points in the mental rotation task. The constant term is 0.219, which meant that the average belief in the score difference was 0.219 points in favor towards men in the base control group. Adding the increase from the update treatment to the constant term comes close to the 1.6 points gap that was reported to subjects. When running the regression results separately for each gender, adding the estimates to the constant term is also close to 1.6 points. Similarly, for the GRE Verbal task, the constant term of -0.5 indicates that in our base control group, subjects believed women outperformed men by 0.5 points. The update treatment, on average, increased the belief in the gap by 0.625 but in favor towards men. Thus, because of the update treatment, our subjects now believe that men performed better by 0.125 points, which is in fact, the score gap reported to subjects.

1.3.2 Do subjects change their preferences as a result of information provision?

Information provision of gender gaps may even influence preferences, causing more men to prefer the mental rotation task, where men performed better, and more women to prefer the GRE Verbal task, where the performance was reported as being about identical. Preference is captured in two ways: Subjects are asked in the questionnaire whether they prefer the mental rotation task or the GRE Verbal task and whether they believed they scored better at the mental rotation task or GRE Verbal task relative to their peers.

We can use the following linear probability model to evaluate whether the updated information affected preferences:

$$Preference^G = \alpha + \delta_1^G Perception + \delta_2^G Update + \delta_3^G Perception * Update + \epsilon_{it} \qquad (1.3.2)$$

Preference is a dummy variable that takes the value of 1 when the mental rotation task is selected and 0 if the GRE Verbal task is selected. It is likely that the updated information will increase the probability that men choose the mental rotation task and decrease the probability that women choose the mental rotation task.

Table 1.4 indicates that there was a shift in preference as well. When subjects are exposed to the update treatment, men increase the probability that men prefer the mental rotation task by 25% (p-value = 0.12) and increase the probability that men claim they are better at the mental rotation task by 31.3% (p-value = 0.026). Conversely, when women were exposed to the update treatment, their probability to prefer the mental rotation decreased by 31.3% (p-value = 0.077) and their probability to believe they performed better in the mental rotation task decreased by 25% (p-value = 0.157). Although two of these results are not statistically significant at the 10% level, given their p-values are close to 0.1 and taken together with the other results, it is still suggestive of a preference shift from the updated information. One interesting result is that the question that yielded the statistically significant estimate for men was framed as a question about confidence, as it asked what task they thought they did better in, while the statistically significant estimate for women was on the question about preferences.

Table 1.4: Task Preferences

Group **Dependent Variable**	**(1)** **Men** **Prefer** **MR**	**(2)** **Men** **Better at** **MR**	**(3)** **Women** **Prefer** **MR**	**(4)** **Women** **Better at** **MR**
Perception Treatment	0.063	0.063	0.063	0.063
	[0.158]	[0.136]	[0.174]	[0.175]
Update Treatment	0.250	0.313**	-0.313*	-0.250
	[0.158]	[0.136]	[0.174]	[0.175]
Perception Treatment*Update Treatment	-0.187	-0.250	0.187	0.187
	[0.223]	[0.192]	[0.246]	[0.247]
constant	0.625***	0.688***	0.625***	0.625***
	[0.112]	[0.096]	[0.123]	[0.124]
N	64	64	64	64
R-Squared	0.04	0.09	0.08	0.06

The post-experiment questionnaire asked subjects if they preferred the mental rotation or the GRE Verbal task and which task they thought they did better relative to their peers. The dependent variable takes a value of 1 if the mental rotation task is selected and 0 for the GRE Verbal task.

1.3.3 Did priming subjects about perceptions and/or information about gender gaps affect their performance?

I use the average score difference from the pre-treatment round and the post-treatment round as the dependent variable, denoted as ΔScore below. The regression is as follows:

$$\Delta Score^{G,T,S} = \alpha + \beta_1^{G,T,S} * Perception + \beta_2^{G,T,S} Update + \beta_3^{G,T,S} Perception * Update + X\gamma + \varepsilon \qquad (1.3.3)$$

S is a superscript that denotes payment style where S = {P, C} for piece-rate or competitive tournament. Perception and Update are the treatment variables described above that take on a value of 1 if treated, 0 otherwise. X is a vector of controls from our post-experiment questionnaire such as race, age, international student status, STEM major, and whether they participate in competitive activities, enjoy competition, believe themselves to be good at competition, and test anxiety. These variables can increase the precision of our estimates.

In the mental rotation task, for the perception treatment, since subjects did not perceive gender differences, I did not expect this treatment to have any effect on future performance. Table 1.5.A shows that this is true in the piece-rate payoff scheme. However, Table 1.5.B shows that under the tournament payoff scheme, there is a statistically significant decrease in the tournament score for the perception treatment for women. This coefficient can be interpreted as an average decrease in 2.322 points when women were asked about their beliefs on what the gender difference is, holding the covariates fixed.

This result is surprising given that women on average, reported there was no gender gap in this task. There are a few possible interpretations of this result. One interpretation is that women were not being completely forthright with their perceptions of gender differences. In addition, it is possible that the subjects were being truthful initially, but the question made the women subjects uneasy as they were uncertain if there was a gender gap. This affected their performance in the tournament stage exclusively, where subjects may be more prone to stereotype threat. However, a more likely explanation is that this result is not robust with a larger sample size. When modifying the regression for our treatment variables to interact with a female dummy variable rather than disaggregating by gender, the estimate of perception treatment interacted with the female dummy is not significant (see Appendix Table B2 and B3).

However, under the information provision treatment, if stereotype threat exists, providing knowledge of an actual gender gap of 1.6 points in favor towards men in the mental rotation task should decrease the scores of women if stereotype threat affects performance. Men may also see an increase in their scores when given information that they outperformed women. However, the information provision treatment had no effect on the score improvement for both genders in the post-treatment rounds in both the piece-rate or the tournament (Table 1.5).[13] The interaction of the perception and update treatments also does not result in any statistically significant effects.

[13] It could be that subjects believe they themselves are "above average", and the 1.6-point difference in the mental rotation task between men and women was not large enough to elicit a credible stereotype threat for women if they believed that they score above the average woman. However, this gap is still quite large; men performed 16.6% better than women and if this study is to be repeated, the larger 2.6 gap from all 128 subjects can be used (which translates to men performed 27.7% higher than women. If this study were to be repeated, I would use the relative percentage terms as a stereotype threat rather than the raw scores because it may lead to a stronger psychological response.

Table 1.5: Mental Rotation: Score Difference from Pre-Treatment & Post-Treatment Rounds

Group	(1) All	(2) All	(3) Men	(4) Men	(5) Women	(6) Women
A. Piece-Rate						
Perception Treatment	0.125	0.219	0.375	0.886	-0.125	0.125
	[0.755]	[0.772]	[1.098]	[1.131]	[1.049]	[1.115]
Update Treatment	-0.094	0.183	-0.500	-0.752	0.313	0.764
	[0.755]	[0.788]	[1.098]	[1.148]	[1.049]	[1.257]
Perception Treatment*Update Treatment	0.000	-0.090	0.750	0.753	-0.750	-1.641
	[1.068]	[1.143]	[1.553]	[1.690]	[1.484]	[1.804]
constant	2.000***	0.872	2.125***	1.855	1.875**	-1.338
	[0.534]	[3.344]	[0.776]	[4.890]	[0.742]	[6.126]
Covariates Included	No	Yes	No	Yes	No	Yes
N	128	128	64	64	64	64
R-Squared	0.001	0.131	0.019	0.253	0.012	0.22
B. Tournament						
Perception Treatment	-0.969	-1.026	-0.125	0.478	-1.812*	-2.322**
	[0.738]	[0.767]	[1.071]	[1.179]	[1.030]	[1.085]
Update Treatment	0.531	0.793	0.938	0.934	0.125	0.366
	[0.738]	[0.783]	[1.071]	[1.197]	[1.030]	[1.224]
Perception Treatment*Update Treatment	0.625	0.117	-0.563	-1.503	1.812	1.280
	[1.043]	[1.135]	[1.514]	[1.762]	[1.457]	[1.755]
constant	2.250***	4.084	2.125***	1.621	2.375***	7.126
	[0.522]	[3.321]	[0.757]	[5.100]	[0.729]	[5.961]
Covariates Included	No	Yes	No	Yes	No	Yes
N	128	128	64	64	64	64
R-Squared	0.035	0.133	0.019	0.146	0.078	0.285

The dependent variable is the score difference from the post-treatment and pre-treatment rounds (either piece-rate or tournament rounds). Perception and Update treatments are dummy variables that take the value of 1 if the treatment is applied. Included covariates when specified are whether one participates in competitive activity, enjoy competition (1-5), good at competition (1-5), how anxious they felt (1-5), age, whether they are an international student, whether they were a STEM major, and race dummy variables.

Table 1.6: GRE Verbal: Score Difference from Pre-Treatment & Post-Treatment Rounds

Group	(1) All	(2) All	(3) Men	(4) Men	(5) Women	(6) Women
A. Piece-Rate						
Perception Treatment	0.031	0.137	-0.063	0.247	0.125	0.238
	[0.400]	[0.423]	[0.567]	[0.588]	[0.582]	[0.643]
Update Treatment	-0.375	-0.281	-0.438	-0.698	-0.313	0.009
	[0.400]	[0.432]	[0.567]	[0.597]	[0.582]	[0.725]
Perception Treatment*Update Treatment	-0.125	-0.324	-0.062	-0.399	-0.187	-0.563
	[0.566]	[0.627]	[0.802]	[0.879]	[0.824]	[1.040]
constant	0.688**	1.585	0.813**	1.723	0.563	1.609
	[0.283]	[1.833]	[0.401]	[2.544]	[0.412]	[3.534]
Covariates Included	No	Yes	No	Yes	No	Yes
N	128	128	64	64	64	64
R-Squared	0.019	0.087	0.023	0.244	0.017	0.161
B. Tournament						
Perception Treatment	0.313	0.391	0.25	0.365	0.375	0.376
	[0.396]	[0.414]	[0.559]	[0.589]	[0.575]	[0.654]
Update Treatment	-0.031	0.182	-0.313	-0.068	0.25	-0.074
	[0.396]	[0.422]	[0.559]	[0.598]	[0.575]	[0.737]
Perception Treatment*Update Treatment	-0.156	-0.530	0.000	-0.614	-0.313	0.008
	[0.560]	[0.612]	[0.790]	[0.881]	[0.813]	[1.057]
constant	0.344	0.746	0.375	-0.033	0.313	1.467
	[0.280]	[1.791]	[0.395]	[2.549]	[0.406]	[3.591]
Covariates Included	No	Yes	No	Yes	No	Yes
N	128	128	64	64	64	64
R-Squared	0.007	0.100	0.017	0.214	0.008	0.102

The dependent variable is the score difference from the post-treatment and pre-treatment rounds (either piece-rate or tournament rounds). Perception and Update treatments are dummy variables that take the value of 1 if the treatment is applied. Included covariates when specified are whether one participates in competitive activity, enjoy competition (1-5), good at competition (1-5), how anxious they felt (1-5), age, whether they are an international student, whether they were a STEM major, and race dummy variables.

For the GRE Verbal task, where subjects perceived that women outperformed men, if men are vulnerable to stereotype threat brought by reminding subjects of these differences in the perception treatment, then they will decrease their performance in the future rounds. Some studies claim even traditionally advantaged groups can also be susceptible to stereotype threat (Aronson et al., 1999).

There should not be an effect on women, because they did not perceive as large of a gender difference as men, and this difference is not very statistically significant. Table 1.6 shows no statistically significant effects for the perception treatment in the GRE Verbal task, where men strongly believed that women performed better, for both men and women.

Also, in the GRE Verbal task, if given the updated information that men and women performed about equally, men may get a boost to their scores if they believed they would have underperformed. However, it also appears that updating men that they perform about equally with women also does not have a statistically significant effect on their score difference from the post-treatment and pre-treatment rounds. If men believed that there was a gender gap in favor of women, providing them information that they performed equally, even marginally better, should have improved their scores.

1.3.4 Are the scores different under piece-rate and competition?

Table 1.7: Pre-Treatment Competition Comparison

A. Comparison of Piece-Rate and Tournament	**Piece-Rate**	**Tournament**	**Difference**	**p-values**	**# men**	**# women**
Men						
Mental Rotation	11.953	11.672	0.281	(0.396)[0.415]	64	0
GRE Verbal	2.156	2.484	-0.328*	(0.064)[0.090]	64	0
Women						
Mental Rotation	9.438	9.063	0.375	(0.328)[0.338]	0	64
GRE Verbal	2.047	1.969	0.078	(0.699)[0.708]	0	64

B. Comparison of Differences by Gender	**Men**	**Women**	**Difference**	**p-values**	**# men**	**# women**
Mental Rotation	0.281	0.375	-0.094	(0.852)[0.823]	64	64
GRE Verbal	-0.328	0.078	-0.406	(0.129)[0.153]	64	64

A. compares the scores of the piece-rate and tournament to determine if these scores are significantly different depending on the presence of competition for each gender. B. compares these differences to see if they are statistically significant between men and women. For the p-values, the parenthesis uses two-sided t tests. The brackets use the Wilcoxon signed-rank test in part A and the Mann-Whitney U test in part B

First, I analyze whether there are differences in outcomes under the piece-rate and tournament in the absence of any stereotype threat treatment. This allows a within-subjects comparison of all 128 subjects. Some early research claims that the gender gap widens under competition (Gneezy et al., 2003) although some more recent experimental evidence has indicated results are ambiguous dependent on the type of task (Shurchkov. 2012; Iriberri and Rey-Biel, 2017). Table 1.7 shows there is no statistically significant difference between the piece-rate and the tournament scores for men and women for the mental rotation tasks. However, in the GRE Verbal task, men performed about 0.328 points better under the tournament than the piece-rate, which was statistically significant. However, a two-sided t test or a Mann-Whitney U test on the score differences between the piece-rate and the tournament is not statistically significant when comparing men and women (Table 1.7 Panel B). Thus, there is no generalizable pattern about whether men or women differ in their outcomes under competition in the absence of the treatment variables.

The next step is to test that whether the stereotype threat treatments will create different outcomes under competition. Table 1.8 summarizes the average score improvements from the pre-treatment to post-treatment rounds in each of the groups, by gender and payment style. A positive value indicates that the subject improved post-treatment while a negative value indicates that a subject performed worse post-treatment. Column (3 and 4) and Column (7 and 8) show non-significant differences in the score improvements between men or women in both the piece-rate and tournament for both types of tasks. Next, I compare the difference between the competition and piece-rate improvement by gender (Columns 9 and 10) to observe whether men and women react differently in each of the groups to different payment styles. Column (11 and 12) shows the difference between the competition improvement and the piece-rate improvement between men and women is not statistically significant. Thus, there is no evidence supporting that men and women react to our stereotype threat treatments differently when exposed to a competitive lab setting. In fact, even in the absence of these treatment, our results do not provide compelling evidence that subjects' performances differ by gender between a competitive and non-competitive payment style.

Finally, the post-experiment questionnaire supports gender differences in attitudes towards competition despite a lack of convincing evidence on performance (Table 1.9). Men are 37.5% more likely to participate in a competitive activity. In addition, on a scale of 1 to 5, men selected a higher self-assessment score on whether they are good at competition and whether they enjoy competition. These differences are highly statistically significant and consistent with the idea that women "shy away from competition" (Niederle and Vesterlund, 2007). However, when asked whether subjects preferred the tournament or the piece-rate, there was no statistically significant difference of their choice by gender, which somewhat contradicts this idea. Nevertheless, it is important to note that Niederle and Vesterlund made their subjects choose a payment style for the final round of tasks as opposed to self-reported preferences.

Table 1.8: Comparison of Score Improvements (Post-Treatment Round - Pre-Treatment Round) by Payment Style

	Piece-Rate Improvement				Competition Improvement				Competition Improve. - Piece-Rate Improve.			
	(1) Men	(2) Women	(3) Diff	(4) p-value	(5) Men	(6) Women	(7) Diff	(8) p-value	(9) Men	(10) Women	(11) Diff	(12) p-value
Mental Rotation												
All Subjects	2.250	1.781	0.469	(0.377)[0.475]	2.391	1.984	0.406	(0.441)[0.599]	0.141	0.203	-0.063	(0.927)[0.750]
Control	2.125	1.875	0.250	(0.778)[0.621]	2.125	2.375	-0.250	(0.775)[0.805]	0.000	0.500	-0.500	(0.634)[0.469]
Perception Treatment	2.500	1.750	0.750	(0.550)[0.531]	2.000	0.563	1.438	(0.242)[0.245]	-0.500	-1.188	-0.844	(0.643)[0.596]
Update Treatment	1.625	2.188	-0.563	(0.599)[0.518]	3.063	2.500	0.563	(0.557)[0.733]	1.438	0.313	1.125	(0.482)[0.448]
Perception and Update Treatments	2.750	1.313	1.438	(0.197)[0.383]	2.375	2.500	-0.125	(0.914)[0.662]	-0.375	1.188	-1.563	(0.251)[0.129]
GRE Verbal												
All Subjects	0.547	0.422	0.125	(0.660)[0.597]	0.344	0.547	-0.203	(0.467)[0.459]	-0.203	0.125	-0.328	(0.424)[0.300]
Control	0.813	0.563	0.250	(0.680)[0.863]	0.375	0.313	0.063	(0.907)[0.787]	-0.438	-0.250	-0.188	(0.831)[0.879]
Perception Treatment	0.750	0.688	0.063	(0.923)[0.788]	0.625	0.688	-0.063	(0.923)[0.909]	-0.125	0.000	-0.125	(0.901)[0.662]
Update Treatment	0.375	0.250	0.125	(0.779)[0.557]	0.063	0.563	-0.500	(0.413)[0.338]	-0.313	0.313	-0.625	(0.343)[0.320]
Perception and Update Treatments	0.250	0.188	0.063	(0.917)[0.862]	0.313	0.625	-0.313	(0.520)[0.406]	0.063	0.438	-0.375	(0.633)[0.370]

Columns (1), (2), (5), and (6) are the average improvement from the post-treatment round and the pre-treatment round by gender. Column (3) and (7) is the difference between the improvements. A positive value indicates that men improved more than women. (9) and (10) are the difference between the competition and piece-rate improvements. A positive value indicates the improvement was larger for the tournament while a negative value indicates a larger improvement for piece-rate. Column (11) is the gender difference, where a postive value means that men had a larger improvement in the tournament. (4), (5), and (12) are the p-values, the parenthesis indicates a two-sided t test while the brackets used a Mann-Whitney U test for non-parametric assumptions.

Table 1.9: Post-Experiment Questionnaire on Competition

	(1) Men	(2) Women	(3) Difference	(4) p-value
Prefer Tournament (0 or 1)	0.438	0.438	0.000	(1)[1]
Participate in Competitive Activity (0 or 1)	0.500	0.125	0.375***	(0.000)[0.000]
Good at Competition (1-5)	3.625	2.969	0.656***	(0.000)[0.000]
Enjoy Competition (1-5)	3.781	3.125	0.656***	(0.000)[0.000]

The parenthesis indicate a two-sided t-test and the brackets are from a Mann-Whitney U test.

1.4 Discussion

While I find convincing evidence that our information provision treatment was credible to subjects by changing their ex-post beliefs in the gender gap, this treatment does not affect post-treatment scores for men or women as one may expect in the presence of stereotype threat. However, the main finding is that information on gender gaps shifted men's preferences towards the mental rotation task, by increasing their belief that they did better in that task relative to the GRE Verbal, and women's preferences towards the GRE Verbal task, by increasing their preference for the GRE Verbal task relative to the mental rotation. Therefore, while stereotype threat may not manifest itself in the form of decreased performance, providing information about group outcomes may have a subtle effect on individual preferences. Assuming external validity, one could imagine that this may be important if stereotypes on group averages dissuade women from certain subjects in school or career fields, even if stereotype threat does not necessarily hinder women's performance in those fields.

Additionally, there is weak evidence that the perception treatment, designed to replicate previous studies in its reliance on implicit beliefs, may have decreased women's performance in the mental rotation tournament round. However, given that women did not actually report that they believed in a gender gap and that this effect is not robust with a modified regression that interacts gender with our treatment variables, it is hard to place too much stock in this finding without a larger sample size. Furthermore, I do not find significant effects with the information provision treatment. Finally, I do not find evidence that men outperform women in competitive settings, which seems to be in line with more recent studies (Shurchkov, 2012; Iriberri and Rey-Biel 2017). I also do not find that men and women react differently in stereotype threat treatments when they are competing in a tournament.

Overall, this experiment adapts previous designs and develops a novel approach to eliciting perceptions of stereotypes and updating information with actual performance outcomes as stereotype treat treatments rather than relying on implicit beliefs or vague statements about group differences. The information provision treatment used in this study uses data from an identical task, with an identical subject pool, and an actual magnitude of the performance gap to generate a credible stereotype threat. Finally, this experiment is one of the first to explicitly test stereotype threat under competitive outcomes. However, while the information is credible due to the corresponding shift in beliefs in the gender gap, it is possible that providing information on magnitudes of the gender gaps has different effects when compared to only providing information about the gender that performed better without information about magnitudes. It could be that even though subjects believed the gender gap that was provided, it was not perceived by the subjects to be very large and if the stereotyped group overestimated their own abilities relative to the group average, they may be less prone to respond to the stereotype threat.

There are potential avenues for future studies. I discussed testing for stereotype threat without the absence of magnitudes as one avenue, which could be done by adding a treatment where the only information provided is that men or women performed better without providing the magnitude of the gender gap. This may prove that stereotype threat is stronger when the threat is subtle. Additionally, it is also worthwhile to consider what would happen in a task with a much larger gender gap, either by using the new pre-treatment scores of the entire 128 subject pool or finding another task with even larger gender differences in favor towards men. [14] Additionally, while this study focused primarily on two tasks that could create stereotype threat for women by updating their perceptions that they did worse on the task than expected, one could think of testing for situations with reducing the stereotype threat. For example, one scenario is if women perceive a large gender gap in favor of men but find out that this gap is small or non-existent. Finally, it might be note-worthy to assess competition between subjects as another treatment variable. This will allow for easier comparisons of its interaction with our treatment variables and provide an overall cleaner analysis. The within-subjects design, although used in earlier competition studies, requires subjects to pay attention and understand the payment structure in each round. Nevertheless, this experiment provides a promising design in evaluating the role of updating information on stereotypes in competitive settings.

[14] Due to IRB protocol forbidding deception, experiments must find tasks that yield actual, large gender differences, which is a costly endeavor. Psychology experiments have more flexibility in deceiving subjects provided they are briefed at the conclusion of the experiment.

2. Relative Sizes of Age Cohorts and Labor Force Participation of Older Workers

2.1 Introduction

In light of population aging in the United States – a development shared with many countries – future employment rates of older individuals will be important determinants of the financial solvency of Social Security, mainly because higher employment implies a continued inflow of Social Security payroll taxes. Illustrating the importance of the employment older individuals, assumptions about labor force participation (LFP) by age play a key role in the 2016 annual report of the federal OASDI and DI trust funds.[15] Moreover, policy responses to population aging seek in one way or another to increase employment of older individuals, such as the increases in the Full Retirement Age and changes in for early retirees reflected in the Social Security Amendments of 1983. The effects of these reforms, and reforms likely to be contemplated in the future, hinge in large part on employment prospects of individuals at older ages.

Our analysis in this paper studies what the changing demographic structure of the United States population implies for the likelihood of employment at older ages. In particular, the overriding question that motivates our analysis is whether increases in the relative shares of the population at older ages are likely to substantially change employment of older individuals. Likely changes in employment independent of Social Security reforms may, for example, lead to increased employment of older individuals that mitigates the increase in the dependency ratio we might otherwise expect from population aging, and condition how we view the anticipated costs or burdens on older individuals from raising the retirement age and reducing benefits for early retirement and Social Security claiming.

The Baby Boom and other, less-dramatic fluctuations in the sizes of birth cohorts generate substantial shifts in the relative sizes of older vs. younger cohorts. Existing work on the effects of cohort size on labor markets in the United States has tended to focus on the effects of own cohort size on wages (e.g., Welch, 1979), and sometimes on employment or unemployment (e.g., Korenman and Neumark, 2000). These studies (as well as work for other countries, such as Morin (2015) for Canada), have tended to focus on the effects on youths of entering the labor market as part of a large cohort. In general, past studies find that youths entering the labor market as part of large cohorts fare

[15] See https://www.ssa.gov/oact/tr/2016/tr2016.pdf, Chapter V.B.5 (viewed April 18, 2017).

worse – earning lower wages, and as a result having lower employment rates – at least initially. These effects are interpreted as "relative supply" or "cohort crowding" effects of a cohort's relative size, with a large cohort shifting out labor supply, depressing wages, and hence lowering employment or labor force participation rates (via the reservation wage effect). The evidence that larger cohorts experience relative earnings declines implies that workers in different age cohorts are only imperfectly substitutable, and some work (e.g., Morin, 2015) suggests – as seems quite plausible – that the degree of substitutability between cohorts is lower the larger the age difference between them.

Our focus in this paper is on older workers – in particular, the effects of the relative size of older cohorts on their labor force participation (LFP) and wages. We concentrate on estimating effects among 50-59 year-olds and 60-69 year-olds. These are the age ranges in which labor force participation first starts to decline, and then when most people retire (see Appendix Table B4). The 60-69 age range, in particular, is the age range in which – in light of population aging – policymakers are trying to increase employment, often through reforms to public pension systems (e.g., Gruber and Wise, 2007). Moreover, this is an age range in which policy may have considerable scope for increasing LFP because of low LFP rates (see Figure 2.1, Panel A).[16]

In the standard relative supply framework applied to younger workers, we would simply view larger older cohorts as likely to experience lower wages and hence lower employment or LFP. Some past work suggests we should not expect much impact of relative cohort size on older workers. For example, Welch (1979) finds evidence suggesting that the adverse effect of entering the job market in a large cohort weakens at older ages, although it does not dissipate. Wright (1991), for the United Kingdom, finds that the effect fully dissipates. However, aside from being quite dated, these studies did not focus explicitly on older individuals. Moreover, if the degree of substitution is quite high between older cohorts and other more-experienced workers, consistent with the flattening of earnings-experience profiles by middle age (Heckman et al., 2006), we might not expect much effect on wages or LFP of being in large cohort of older workers.

Despite these considerations, there are reasons to expect that the effects of cohort size could be sizable for older workers. Older individuals in their 50s or 60s have low employment rates relative to those in their 40s or 30s – in part because of transitions to retirement, especially in the 60s (e.g., Munnell, 2015). At the same time, retirement is quite fluid, because many seniors transition to part-

[16] We explored grouping the 50-59 year-olds with 25-49 year-olds, but the data indicated that, for the analyses we present, the behavior of 50-59 year-olds was similar to that of 60-69 year-olds, and dissimilar to that of 25-49 year-olds.

time or shorter-term "partial retirement" or "bridge jobs" at the end of their careers (e.g., Johnson et al., 2009), or return to work after a period of retirement (Maestas, 2010). Together, these facts suggest that older workers may have quite elastic labor supply on the extensive margin, in contrast to workers (especially men) of other ages, in which case the effects of large cohort size on LFP or employment, stemming from wage effects, could be sizable.[17] Moreover, if older workers in partial retirement are leaving career jobs, and perhaps taking lower-skilled or less-demanding jobs, they may not be so substitutable with prime-age workers, implying that there could be larger effects of cohort size on wages in this age range – like for young labor market entrants.

To this point, we have focused on the usual relative supply hypothesis about cohort size, which predicts negative effects of large relative cohorts on LFP and wages. However, there are two factors that could push in the opposite direction, towards a positive effect. First, we might expect the age structure of the population to affect the composition of consumption and hence labor demand.[18] It is possible that the age structure of employment is such that relative labor demand for an age cohort increases when the relative size of that cohort increases. For an example particularly pertinent to older workers, Cohen (2006) documents the aging of the U.S. nursing workforce, for which demand will surely grow as the population ages.

Second, a relative cohort size measure is just that – a relative measure. Thus, an increase, say, in the size of the 60-69 year-old cohort relative to the population means that the old cohort is large relative to at least some other narrowly-defined age cohorts. If two age cohorts are substitutable, then a decline in the relative size of one of them can imply an increase in the relative demand for the other. For example – again with particularly relevance to older workers – the partial/bridge retirement phenomenon may mean that "post-retirement" workers take lower-skilled jobs more similar to those held by younger workers, in which case older workers could be substitutable with young workers and a large cohort of 60-69 year-olds relative to young workers can increase demand for 60-69 year-olds. Alternatively, if older workers are more substitutable for workers in the prime/middle-aged cohort, we might find this positive demand response for the size of the older cohort relative to this cohort.

We explore the effects of the relative sizes of age cohorts on LFP and wages, focusing on the effects on older individuals. We use long-term data on cohort size and cohort labor force participation

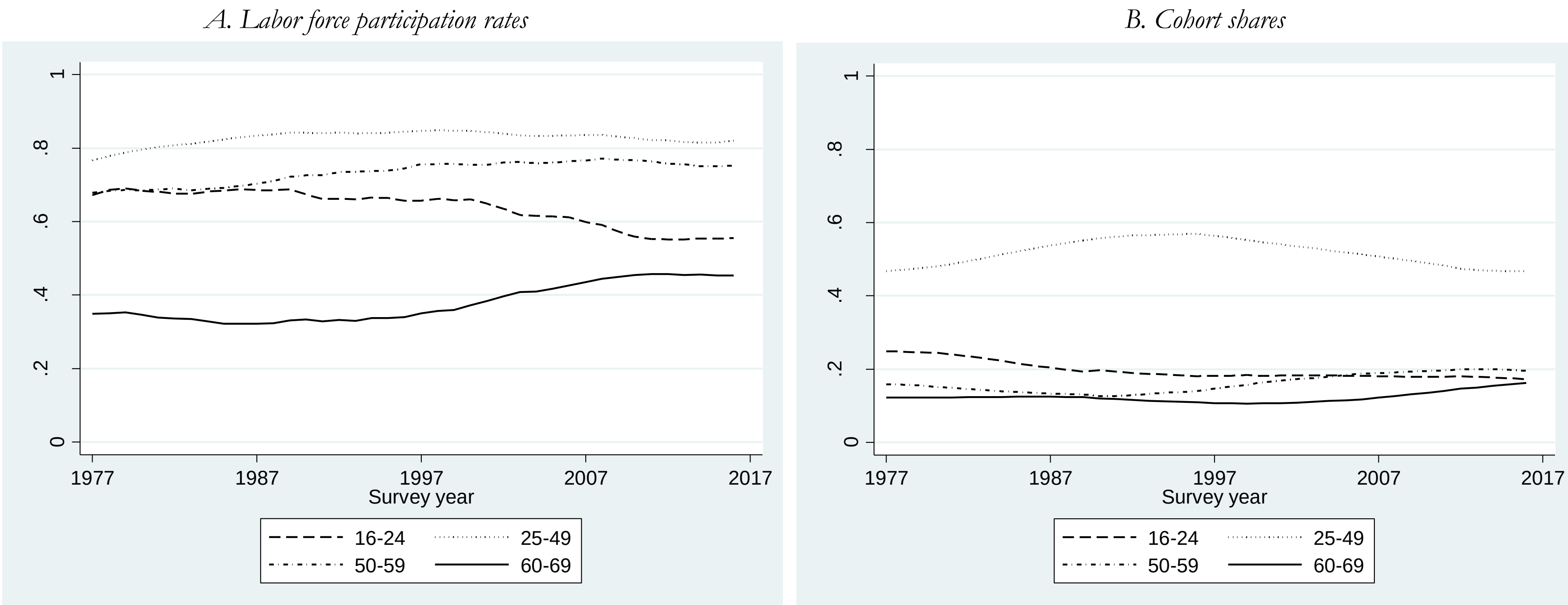
---

[17] The meta-analysis in Evers et al. (2008) points to a very low extensive margin labor supply elasticity for men generally. For evidence suggesting sizable extensive margin labor supply elasticities for older workers in the United States, see French and Jones (2012).

[18] For example, Reinhardt (2003, Exhibit 1) reports that per capita health spending for 55-64 year-olds is double that for 25-34 year-olds.

rates and wages over many decades, exploiting variation across states in a panel data setting that controls for other influences on employment of older workers, as well as for some measures related to longer-term, life-cycle responses to cohort size, like education and marriage. We pay careful attention to the endogeneity of the contemporaneous age structure of a state's potential workforce. Given distinct and persistent patterns of internal migration related to age (e.g., migration to Florida and Arizona), as well as more variable changes in internal migration with respect to economic conditions (and international immigration), we might expect the effects of the relative sizes of different age cohorts to be hard to detect in OLS estimates. For example, an adverse effect of a large cohort on LFP may be obscured because the cohort is large owing to in-migration in response to strong labor demand. We instrument for contemporaneous relative cohort size measures using historical birth data by state and cohort – which should be an exogenous source of variation in states' current demographic structures.

Figure 2.1: Labor Force Participation Rates and Population Share by Age Group, Over Time



Source: Census Population Survey (CPS) 1977-2016. In Panel A, a state panel is first constructed from CPS monthly basic files by aggregating labor force participation for each state, year, and age group. The figure in the panel is created from weighted averages of all states' labor force participation rates, weighted by state population. In Panel B, cohort share is constructed from the CPS monthly basic files by dividing the sum of the CPS survey weights for each age group in each state and year by the total sum of the survey weights for ages 16-69 in each state and year. The figure in the panel is constructed from weighted averages of all the states' cohort shares, weighted by state population.

.

2.2 Relevant Prior Work

There is long-standing interest in factors affecting the employment of older workers, often motivated by implications for retirement systems. Perhaps the largest body of research focuses on work incentives created by the Social Security system itself – including the level of benefits (e.g., Burtless, 1986), the early retirement age (e.g., Gustman and Steinmeier, 2005), the structure of the earnings test (e.g., Friedberg, 2000), and the impact of reforms to delay retirement (e.g., Neumark and Song, 2013).

Research has also focused on other factors affecting employment of older workers. For example, there has been an outpouring of research on factors that appeared to have slowed the growth in employment and labor force participation of older workers since the Great Recession, such as changes in age discrimination (Neumark and Button, 2014) and increases in SSDI awards (Mueller et al., 2016).

The effect of the relative sizes of age cohorts on the LFP of older individuals is a potentially important factor to study, for at least two reasons. First, variation in cohort size can be used to improve predictions of long-run changes, because the sizes of age cohorts can be quite reliably projected far into the future.

Second, past research on the effects of cohort size on young workers establishes that cohort size can be influential. Welch (1979) showed that within schooling groups, the large cohort size of Baby Boomers reduced wages, with a larger impact on highly-educated workers and workers early in their career.[19,20] Korenman and Neumark (2000) study variation over countries and across time to estimate the effect of the relative size of youth cohorts on youth unemployment rates. Like the strategy we use in this paper, they use an instrumental variables approach based on births by cohort and country to account for the endogeneity of cohort size with respect to labor market conditions (via migration), and when doing so find that larger youth cohorts are associated with higher unemployment rates.

---

[19] Welch's study, like many others on cohort size, focuses on wages, but the effects of cohort size on wages should translate into effects on employment and LFP rates, with lower wages reducing these rates, and vice versa. Berger (1984) builds on Welch's work by looking at effects on earnings profiles, which he interprets as reflecting human capital investment. This channel of influence is less relevant for older workers.

[20] Macunovich (1999) tried to separate labor supply and labor demand effects of cohort size, suggesting that relative sizes of birth cohorts (and changes in sizes of birth cohorts, to capture leading and lagging effects of a boom) affect supply, while relative sizes of current cohorts (and changes) reflect demand. It is not clear why this distinction isolates supply and demand effects; indeed, we use data on births to construct instrumental variables for contemporaneous cohort sizes, without taking a position on whether births drive supply or demand.

We do not focus only on the relative cohort size of the cohort of interest – in this case, older individuals – but also on a more detailed characterization of the sizes of other cohorts in different age ranges. This can matter because substitutability between cohorts may vary with "distance" in age. One paper that pays more attention to sizes of multiple cohorts is Stapleton and Young (1988), although they focus more on incentives to invest in education owing to how substitutability between cohorts varies by education, a question farther removed from the focus of our paper. Our research also differs in focusing on how cohort size affects LFP (and wages) of older individuals.

2.3 Empirical Specifications and Strategy

We begin with a standard relative cohort size specification used to estimate the effects of a large cohort of older individuals on their LFP. This specification takes the form:

$$LFP_{st}^{O} = \alpha + \beta^{O/T}\mathrm{RCS}_{st}^{O/T} + X_{st}\gamma + \lambda_s + \theta_t + \epsilon_{st} \qquad (2.3.1)$$

The O superscript denotes older cohorts aged either 50-59 or 60-69. RCS is a relative cohort size measure, and the O/T superscript denotes this that is computed for older cohorts relative to all working-age cohorts (16-69).[21] X is vector of controls including: the unemployment rate for 16-49 year-olds,[22] the rate of state GDP growth from the previous year to the current one; the shares married and living together, divorced or widowed, and spouse absent or separated;[23] the shares female (when we estimate regressions for men and women combined), Hispanic, black, urban, and union members; and the shares with less than a high school degree and a bachelor's degree or higher. The marital status and education controls may reflect some of the life-cycle responses to variation in cohort size that reflect decisions taken at earlier ages, but also other changes that have a large exogenous component; with the CPS data we use, we are unable to measure years spent in different marital status states. The s and t subscripts denote state and year, and λ_s and θ_t are vectors of fixed state and year effects.

The year effects we can incorporate into the state-level panel data we study are potentially very important. There are many economy-wide changes, including factors such as technology, trade liberalization, increases in women's LFP, and declines in marriage, which can influence employment of individuals of different ages. In aggregate data, there would be no way to control for these

[21] We verified that estimating equation (2.3.1) for the size of the 60-69 cohort or the 50-59 cohort relative to the 16-49 cohort (or the 16-59 cohort, for the analysis of 60-69 year-olds) yields very similar results to defining the size of the older cohorts relative to 16-69 year-olds.
[22] This is defined for men, women, or both sexes, corresponding to the sample used in the regression.
[23] Until 1989, the data combined divorced and widowed, and combined spouse absent and separated.

potentially confounding influences.

LFP is the state-by-year average. The LFP and RCS variables are entered in logs, so $\beta^{O/T}$ is an elasticity. Because we use sample estimates of state-level averages to construct our data, we always use GLS, weighting by average state population measured over the sample period. Our IV estimates (described below) are similarly weighted. We also estimate versions of equation (2.3.1) for the state-by-year log of average hourly wages.[24]

The estimate of $\beta^{O/T}$ measures the impact of the size of the older cohort relative to the workforce on LFP (or wages) of that older cohort. We would expect similar qualitative results for cohorts of other ages, viewed through the simple mechanism of supply shifts. We hence also estimate equation (2.3.1) for younger cohorts (aged 16-24, denoting the cohort size variable $\mathrm{RCS}^{Y/T}$), and prime-aged cohorts (aged 25-49, denoting the cohort size variable $\mathrm{RCS}^{P/T}$), with the corresponding coefficients defined as $\beta^{Y/T}$ and $\beta^{P/T}$).

The relative cohort size measures may be endogenous. One possibility is that people migrate to where labor market conditions for their age group are better. This would create a bias against finding evidence, predicted by the relative supply hypothesis, that a larger relative cohort size reduces LFP or wages, as the cohort size may expand in response to high labor demand (which boosts LFP and wages). We might expect this kind of migration to be more common for younger cohorts.

In contrast, older individuals may be more likely to migrate for retirement-related reasons. States that are retirement destinations will tend to have larger relative older cohort sizes but lower LFP rates – not because of cohort-size effects on labor supply, but through selective in-migration of older retirees. And similarly, states *from* which retirees (or near-retirees) migrate will tend to have lower relative cohort sizes at older ages, but high LFP, because of selective out-migration of retirees. The endogeneity bias from retirement-induced migration is thus in the opposite direction to the endogeneity bias from employment-induced migration – with retirement-induced migration biasing the evidence in favor of evidence, when looking at LFP, for the relative supply hypothesis. Of course, it is possible that some older people migrate based on labor market conditions, if they entertain the possibility of some "post-retirement" work, so the direction of bias is ultimately an empirical question. Finally, in contrast to LFP, there is no clear prediction about bias in the estimates of equation (2.3.1)

[24] We estimated all of our main specifications without weighting (available upon request). We put less store in the unweighted estimates, because the first-stage is much weaker, likely because of the greater weight put on smaller states with less accurate estimates. (Correspondingly, the reduced-form relationships are also much weaker.) That said, if we compare the OLS estimates, there is no indication of substantive differences in the results.

for wages – for older cohorts – from retirement-related migration.

Migration flows appear to be large enough to matter, and this is borne out in our instrumental variables estimates. Appendix Figures A1-A4 show data on interstate in-migration rates for retirement-related and work-related reasons, based on CPS Annual Social and Economic Supplement (ASEC) data. Appendix Figure A1 shows data for 60-69 year-olds, with states ordered by retirement related in-migration rates. The states near the top of this list – like Arizona, Florida, and Nevada – are not surprising. For these states, the one-year in-migration rates are near 0.4 percent. Thus, interstate in-migration could, over a number of years, result in sizable changes in the cohort share. We see, by the way, a good deal of work-related interstate migration reported for this age group, suggesting that it is unclear what the direction of bias might be when we estimate equation (2.3.1) for the older cohort; this depends on both the magnitudes of the migration flows, as well as their endogeneity.

Appendix Figure A2 shows the data for 50-59 year-olds. There is less retirement-related migration for this age group; for the states with the highest rates, the level is about half (0.2 percentage points) what it is for 60-69 year-olds. Interestingly, also, the states for which retirement-related migration is highest are somewhat different than for 60-69 year-olds. For 50-59 year-olds, far more migration is work related.

Appendix Figures A3 and A4 show the data for the other two, younger cohorts, now with states ordered by work-related in-migration rates. There is, not surprisingly, very little retirement-related migration for these age cohorts. But work-related in-migration rates are often quite high, with one-year rates well above one percent for 25-59 year-olds, and 1.5 percent for 16-24 year-olds. Thus, again, over many years in-migration could have substantial effects on the cohort share.

To address the potential endogeneity of relative cohort size, we instrument for the relative cohort size variables using predicted relative cohort sizes based on past births in the state for the years in which members of a cohort would have been born. Thus, for example, the instrumental variable for $\text{RCS}^{O/T}$ in 2000 – the ratio, in 2000, of the number of people currently in the state aged 60-69, divided by the number aged 16-69 – is the ratio of the number of people born in the state between 1931 and 1940, to the number of people born in the state between 1931 and 1984. The logic of this instrumental variable is clear. The relative birth-cohort size instrument should predict the contemporaneous relative cohort size quite well – and it does. And it is hard to fathom a reason why the relative birth-cohort size instrument – often constructed from very long lags – would affect current labor market outcomes conditional on the contemporaneous relative cohort size variable; hence, the instrument should satisfy the exclusion restriction.[25] Thus, the relative birth-cohort instrument should purge the contemporaneous relative cohort size variable of variation attributable to migration. (It should also help correct for other sources of bias, such as measurement error in the estimation of the contemporaneous relative birth cohort variables; the latter are estimated from the CPS, whereas the birth cohort variables are constructed from the universe of birth records.)

The standard expectation, based on the relative supply hypothesis regarding cohort size, is that the effects of $\text{RCS}^{O/T}$ on LFP and wages will be negative, and similarly for $\text{RCS}^{Y/T}$ and $\text{RCS}^{P/T}$ when we look at the younger cohorts. However, if older cohorts have more elastic extensive margin labor supply responses, we might find larger negative estimates for LFP of older cohorts. In contrast, the effects of relative cohort size could go in the other direction because of effects of age structure on the age composition of labor demand, and other differences between cohorts could arise because of substitution between workers in different age cohorts.

Our main analysis extends beyond equation (2.3.1). In particular, we explore whether the effects of age structure on LFP and wages of older workers are more complex than simply an effect of their cohort size relative to the working age population, owing to more complex spillovers between cohorts of different ages. These complexities could arise through the demand side, depending on how the relative sizes of other cohorts affects demand for older workers. They could also arise through the supply side, as a large relative cohort of older workers could be driven by a smaller cohort of very young workers, or of prime/middle-aged workers, and there may be different degrees of substitutability between these cohorts and older workers.

To address these questions, we modify equation (2.3.1) and instead estimate a model with separate effects of the size of the older cohort relative to the two younger age cohorts:

$$LFP_{st}^{O} = \alpha + \beta^{O/Y}\text{RCS}_{st}^{O/Y} + \beta^{O/P}\text{RCS}_{st}^{O/P} + X_{st}\gamma + \lambda_s + \theta_t + \epsilon_{st} \quad (2.3.2)$$

[26]

The estimate of $\beta^{O/Y}$ captures the effect of the size of the older cohort relative to the younger cohort, and the estimate of $\beta^{O/P}$ captures the effect of the size of the older cohort relative to the prime-aged

[25] As indirect evidence, we verified that our instrumental variable does not predict contemporaneous state GDP growth, whether or not we condition on the contemporaneous relative cohort size variable. This holds true across age groups, and in the richer specifications described below with two relative cohort size variables and two instrumental variables. (Results available upon request.)

[26] Because results for equation (2.3.1) were very similar for the size of the 60-69 cohort or the 50-59 cohort relative to the 16-49 (or 16-59) cohort, the differences in results we find using equation (2.3.2) have to do only with differences in the sizes of the older cohorts relative to the sizes of the 16-24 or 25-49 cohorts.

cohort.[27] Equation (2.3.2) can tell us, for example, whether the effect of a large older cohort on LFP varies with whether the older cohort is large relative to the cohort of workers distant in age (i.e., the young), or relative to the cohort of those closer in age. As for equation (2.3.1), we also estimate versions of equation (2.3.2) for the state-by-year log of average hourly wages.[28]

We also address endogeneity bias in equation (2.3.2). Indeed, differential responsiveness of migration across age groups could be particularly problematic in estimating equation (2.3.2). For example, suppose there is strong retirement-related migration of older individuals. We would not expect any such response among the younger cohort; in contrast, there could be at least some retirement-related migration in the prime-age group. In that case, the negative correlation between ε' and $\mathrm{RCS}^{O/Y}$ in equation (2.3.2) could be particularly strong. We use the same overall strategy, but now using two instrumental variables for the two relative cohort size variables in equation (2.3.2). For example, the instrumental variable for the $\mathrm{RCS}^{O/Y}$ in 2000 – the ratio of the number of people currently in the state aged 60-69 in 2000, divided by the number aged 16-24 – is the ratio of the number of people born in the state between 1931 and 1940, relative to the number of people born in the state between 1976 and 1984. And the instrumental variable for the $\mathrm{RCS}^{O/P}$ in 2000 – the ratio of the number of people currently in the state aged 60-69 in 2000, divided by the number aged 25-59 – is the ratio of the number of people born in the state between 1931 and 1940, relative to the number of people born in the state between 1941 and 1975.

2.4 Data

Our contemporaneous population and LFP data come from the Current Population Survey (CPS)

[27] From here, we use "prime" to refer to ages 25-49. This is not meant to reflect a judgement about age. But use of "middle-aged" for 25-49 year-olds is likely to create more confusion.

[28] In research explaining wage differences between groups, a similar specification is sometimes estimated for relative wages. For example, Card and Lemieux (2001) assume a production function including constant elasticity of substitution (CES) subaggregates, by education, of workers of different ages. The analyses in Stapleton and Young (1988) and Welch (1979) have a similar flavor. In contrast to understanding the evolution of wage gaps differences between groups, our primary focus is on understanding the influence of relative cohort size on LFP, and to understand this influence, the effects of relative cohort size on wage levels are most important. Moreover, more recent work on cohort size has focused on the effects on wage levels (e.g., Berger, 2004; Korenman and Neumark, 2000; Macunovich, 1999; and Morin, 2015). In addition, the relative outcome specifications would only make sense for the estimates of equation (2.3.1), and not equation (2.3.2), because the latter includes two different relative cohort size variables on the right-hand side. For these reasons, and because equation (2.3.2) provides our key results (which, regardless, have to do more with effects on LFP than on wages), we focus in the paper on specifications of the effect of relative cohort size on wage levels. However, for completeness, we have estimated the log average wage models corresponding to equation (2.3.1) using relative measures instead (available upon request).

monthly basic files, from 1977-2016.[29] The micro-data are aggregated to create state-by-year measures. Cohort sizes are constructed by weighting individuals by the survey weights used to aggregate up to population estimates, to make the estimates population representative. For example, for the oldest cohort of 60-69 year-olds, $\mathrm{RCS}_{st}^{O/T}$ is constructed by taking the sum of the survey weights in state s, at time t, for ages 60-69 divided by the sum of the survey weights in state s, at time t, for the entire 16-69 age group. LFP rates are constructed using the same survey weights.

Wage data come from the CPS merged outgoing rotation group (ORG) files, which are available from 1979. The hourly wage is measured directly as earnings per hour when available (for those paid hourly). Otherwise, it is constructed by dividing earnings per week by the usual hours worked, for those who are not paid by the hour.[30] The computed hourly wages are trimmed by removing hourly wages below half the state minimum wage or above $200/hour (in 2016 dollars). Hourly wages are then averaged by state and year, using the survey weights.

The instrumental variables construction was considerably more involved. We use historical series on births by states, based on U.S. Vital Statistics reports published by the National Center for Health Statistics (NCHS). The data are available in two forms, either through Births: Final Data reports (NCHS 1974–1975, 1976–1977, 1978–1980, 1982–1996, 1997–1998, 1999–2017) retrievable online from 1970,[31] or in U.S. vital statistics reports (Grove and Hetzel 1968; Linder and Grove 1947; NCHS 1934, 1934–1938, 1939–2005, 1984); both reports are typically published two years after the reported year. The U.S. vital statistics reports have been produced since 1890, although birth information was not captured until 1915, when 10 states and the District of Columbia adopted the birth-registration system (Linder and Grove 1968). Other states began to trickle in, with the final participating states of Texas joining in 1933 and then Alaska joining in 1945 (see Appendix Table B5). The Births: Final Data series is more recent. It started in 1971 and was published concurrently with the U.S. vital statistics reports, but the latter was phased out by 2003. The two reports are not completely identical but do not have large discrepancies.[32] We use the reported numbers of births in the Births: Final Data reports as our source back to and including 1971, and we use the U.S. vital statistics reports for prior

[29] Our data come from the Integrated Public Use Microdata Series (Flood et al., 2017)
[30] Observations are not used if earnings are not reported, or if only weekly earnings, but not hours, are reported.
[31] Births: Final Data is the name of this report in later years. In earlier years the name varied, as reflected in the NCHS documents for the corresponding publication dates listed in the References.
[32] This is based on personal communications with Michelle Osterman, a health statistician at NCHS (5/2/17 and 5/8/17). The Births: Final Data series is easier to navigate and seems to be cited more often.

years back to 1931.[33] Prior to 1931, the number of births is not available, so we reconstruct the level from the crude birth rates, defined as the number of births per 1,000 population.[34]

There is surely some measurement error in the birth instruments we construct. And the accuracy of reporting is worse in the earlier data. For example, our constructed number of births from 1915 to 1930 from crude birth rates and estimated population sizes suggests a sharp decrease in the number of births from 1930 to 1931, which implies we overstated the number of births from 1915 to 1930. This is likely because crude birth rates are inconsistent due to unclear adjustments for under-registration.[35] Overall, the general issues with crude birth rates contributed to our decision to use the number of births from the individual yearly files either from *Births: Final Data* or *U.S. Vital Statistics Reports*, whenever available.

Despite these concerns, measurement error in instrumental variables is of less concern than measurement error in the variables of interest. Indeed, if the measurement error in the instrument is uncorrelated with the variable(s) for which we are instrumenting, and uncorrelated with the error term in the equation of interest, the measurement error does not introduce any inconsistency in the instrumental variables estimation, although it can weaken the instrument and make the instrumental variables estimate less precise. This is true even if the measurement error is worse in earlier periods (i.e., heteroskedastic). Therefore, while we note these potential issues with the early birth data, we do not believe these issues pose substantive challenges to our empirical analysis.

To have data on the birth instrument for the oldest people in our sample (age 69), we shorten the CPS panel we use to begin in 1984, rather than 1977 (for LFP) or 1979 (for wages). Even then, our panel with the instrument is unbalanced because we do not have the requisite birth data for all states from the earliest year, due to when states started reporting births. However, there are no gaps between years. For example, in 1984, there will be ten states and D.C. available since the number of births in the old cohort is drawn from the number of births in years 1915-1924. In later years, more

[33] One exception is the year 1979, in which the final report is not available online. For this year, we use the *U.S. Vital Statistics Report* birth numbers.

[34] *Vital Statistics Rates in the United States 1900-1940* (Linder and Grove 1947) contains birth rates and estimated population sizes for 1915 to 1930, which allows us to estimate the number of births

[35] For example, the crude birth rates from 1915 to 1929 do not have birth rates adjusted for under-registration while the 1930-1940 crude birth rates had an adjustment for under-registration. In general, these earlier adjustments to crude birth rates are not well documented or transparent other than the dates of the adjustments. The most egregious example is that the crude birth rates recorded for 1940 are different between the vital statistics for 1900-1940 (Linder and Grove 1947) and 1940-1960 (Grove and Hetzel 1968). Michelle Osterman and her colleague, Brady Hamilton were unable to reconcile this difference, but believe the more recent statistics (for 1940-1960) is accurate (personal communication, 5/8/17).

states are added as their number of births are reported. For example, Georgia, which first started collecting birth data in 1928, will be first be available in 1997, when the number of births for 69-year-olds is recorded.

2.5 Results

2.5.1 Descriptive Statistics

In Figure 2.1, Panel A shows LFP rates by age group, and Panel B shows cohort population shares. (Appendix Figures A5 and A6 show the corresponding information for wages.) We do not want to infer much from these aggregate time series,[36] but we see that the rise in LFP of 60-69 year-olds in the latter part of the sample period (Figure 2.1, Panel A) coincides with an increase in their relative cohort size (Figure 2.1, Panel B).[37] On the surface, this is inconsistent with the usual relative supply cohort size hypothesis in which a large cohort size depresses LFP. Moreover, Appendix Figure A5 shows that the rising LFP of the older cohort was accompanied by rising real wages,[38] also inconsistent with the relative supply hypothesis.

We next explore the relationships between LFP, wages, and relative cohort size in more detail, providing similar evidence for different age cohorts, and showing both the time-series and the within-state variation and covariation between these three variables. First, to avoid having to compare across the panels in Figure 2.1, Figure 2.2 graphs the time series on LFP rates and relative cohort size for each of the four age cohorts. Panel A of Figure 2.2 – for 16-24 year-olds – parallels the evidence for 60-69 year-olds in that LFP rates and relative cohort size tend to move in the same direction, rather than the opposite direction as predicted by the relative supply hypothesis. The evidence for 25-49 and 50-59 year-olds is less clear.

Appendix Figure A6 shows the same type of evidence, but for real wages. Here, the evidence for the younger cohorts is mixed. The evidence for 25-49 year-olds shows rising wages in the latter part of the sample period, when relative cohort size is declining – consistent with the relative supply effect of cohort size. However, in the earlier part of the sample, wages are flat as relative cohort size rises. And the correlation is negative, as reported in the notes to the figure. For 16-24 year-olds, in

---

[36] However, one potential advantage of the national time-series data, relative to more disaggregated data, is that it should not be influenced by bias from endogenous migration across states. On the other hand, with national time-series data we cannot control for aggregate trends or changes in age-specific labor demand, which could bias our results. It is possible that these trends/changes underlie some of the apparent inconsistencies in the aggregate data.

[37] The correlation is 0.595.

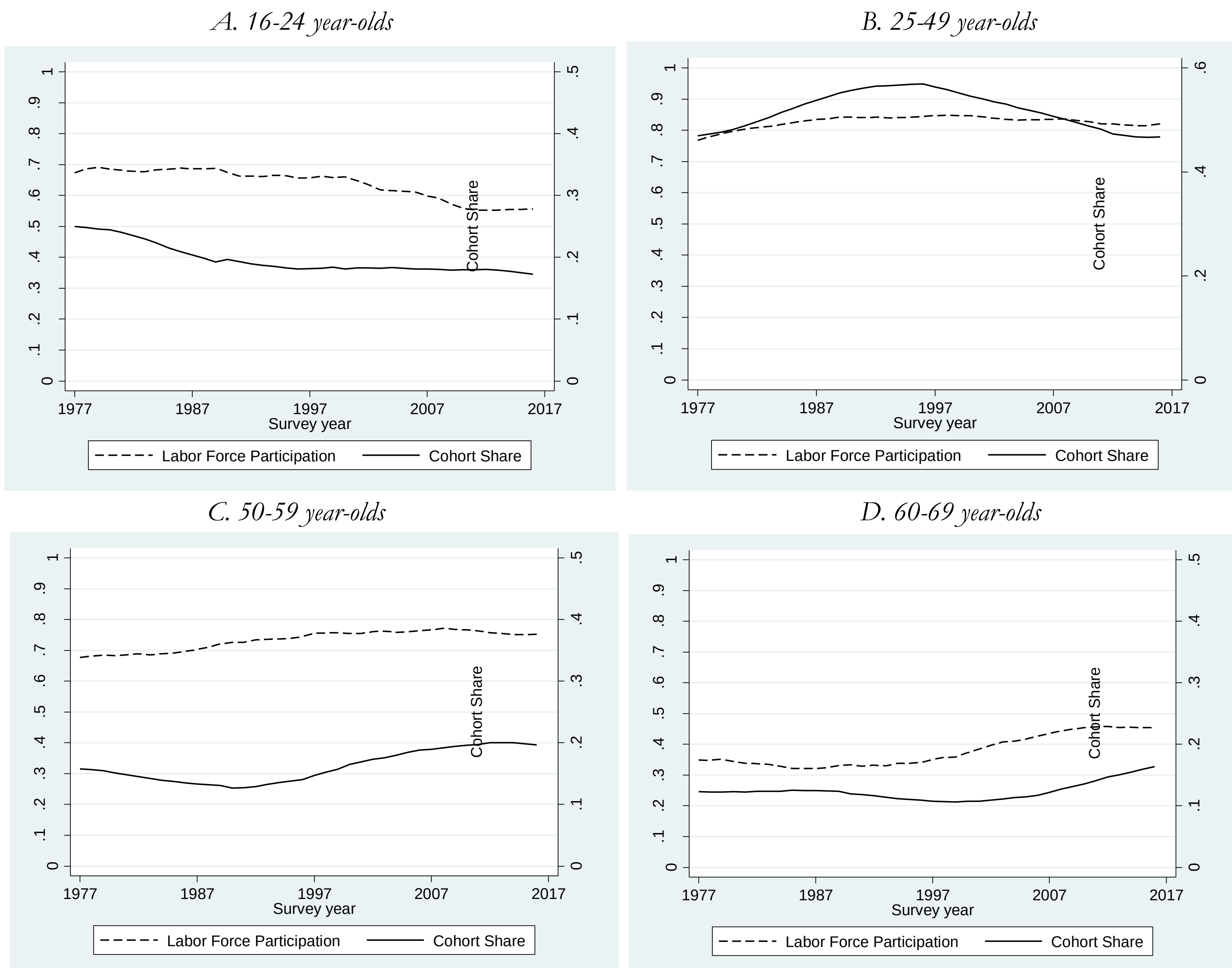
[38] The correlation is 0.627.

contrast, the wage and relative cohort size series track each other in the early part of the sample, which is inconsistent with the relative supply effect of a larger cohort, and then both series are largely flat subsequently. The correlations for this age group, as well as the two older cohorts, are positive; see the figure notes.

The final two appendix figures (A7 and A8) instead provide information on changes over time at the state level, providing scatter plots of the 1977 to 2016 changes (1979 to 2016 for wages) for each state. Thus, the data points summarize the overall changes over the sample period, in contrast to the year-by-year changes graphed for the aggregate time series. In Appendix Figure A7, for LFP rates and relative cohort size, there is evidence of negative relationships for all four age cohorts – 16-24, 25-49, 50-59, and 60-69 – although the slope coefficient is particularly large for 16-24 year-olds (−0.979) and near zero for 50-59 year-olds. (The correlation is only statistically significantly different from zero for 16-24 year-olds.) These contrast with the positive correlations in the time-series data shown in Figure 2.2 and are more consistent with the relative supply effect of cohort size. In Appendix Figure A8, we find evidence of a positive relationship for wages for 16-24, 25-49, and 50-59 year-olds, inconsistent with the relative supply effect of cohort size, while the evidence for 60-69 year-olds is more consistent with this effect.

Thus, the time-series evidence is largely inconsistent with the relative supply effect of cohort size (Figure 2.2 and Appendix Figure A6). The state-level evidence for LFP is consistent with this effect, for all age cohorts (Appendix Figure A7), while in three out of four cases, the evidence for wages is not (Appendix Figure A8). However, this evidence is suggestive at best, and the state-level evidence may be particularly prone to endogeneity bias, with the bias for the older cohorts, for LFP, likely in the negative direction. Hence, we next turn to the regression estimates, with the instrumental variables (IV) estimates most likely to uncover the true effects of relative cohort size.

Figure 2.2: Labor Force Participation Rates and Cohort Shares by Age Group, Over Time



Source: Data source and series construction are explained in notes to Figure 2.1. Pearson correlation coefficients for 16-24, 25-59, and 60-69 year-olds are 0.640, 0.784, 0.628, and 0.595 respectively.

Table 2.1: OLS Regressions of Log Labor Force Participation Rate on Log Cohort Share, 1977-2016

Sex	Both sexes				Men				Women			
Age	**16-24**	**25-49**	**50-59**	**60-69**	**16-24**	**25-49**	**50-59**	**60-69**	**16-24**	**25-49**	**50-59**	**60-69**
ln(Cohort Size/Pop 16-69)	0.101***	0.042	0.046**	-0.131**	0.033	-0.002	0.032*	-0.077	0.101**	0.039	0.062*	-0.099
	[0.036]	[0.032]	[0.021]	[0.063]	[0.033]	[0.016]	[0.017]	[0.059]	[0.038]	[0.057]	[0.035]	[0.082]
Mean LFP	0.64	0.83	0.73	0.38	0.67	0.93	0.83	0.45	0.61	0.73	0.64	0.31
R^2	0.92	0.91	0.90	0.89	0.92	0.91	0.81	0.80	0.89	0.93	0.93	0.88
Incremental R^2	0.029	0.033	0.008	0.011	0.013	0.015	0.024	0.011	0.052	0.037	0.006	0.006

Notes: Data sources are described in notes to Figure 2.1. The table reports estimates of equation (2.3.1). All specifications include fixed state and year effects, controls for the 16-49 unemployment rate, state GDP growth, a constant, and the shares married and live together, share divorced or widowed, share spouse absent or separated, Hispanic, non-Hispanic black, with a bachelor's degree or more, high school dropouts, and urban. The "Both" columns also include the share female. Regression weighted by average state population through the sample period. Standard errors are clustered by state. Estimates are weighted by average state population through the sample period. *, **, and *** denotes 0.1, 0.05, and 0.01 significance. Incremental R^2 refers to addition to R^2 beyond fixed state and year effects. N = 2,040.

2.5.2 LFP: OLS regression estimates

Table 2.1 reports OLS regression estimates of the effects of relative cohort size on LFP, for each age group; these are estimates of equation (2.3.1). We present results for both sexes combined and then for men and women separately.

Over the sample period and ages considered there are potential reasons to prefer one approach or the other. For the older cohorts, LFP and careers of men and women were quite different, and men and women were likely not viewed as highly substitutable labor inputs. The younger cohorts, in contrast, experienced rising LFP of married women, and some convergence in the occupational distribution. Unlike, say, in the research literature on the impacts of relative supplies of workers by educational level, we are unable to disentangle the effects of variation in male and female relative cohort size on outcomes for men and women, because the relative cohort size variables are so highly correlated (over 0.99 partial correlations, conditioning on the other control variables in our models). Thus, we simply report the results separately by sex, and combined, and note that the findings are generally robust.

For both sexes combined, we find a positive and significant effect of relative cohort size for 16-24 year-olds, with an elasticity of 0.101. The estimate for 25-49 year-olds is also positive, but insignificant and smaller (an elasticity of 0.042). The estimated elasticity for 50-59 year-olds is a bit larger (an elasticity of 0.046) and statistically significant. For the oldest cohort, aged 60-69, the estimate is significant and negative, with a larger absolute estimated elasticity (−0.131). The sign pattern of the estimates is almost always the same for men and women separately (with one minor exception for an estimate very close to zero for 25-49 year-old men). In addition, some of the estimates for men and women separately are smaller in absolute value than the estimates for both sexes combined, and combined with larger standard errors, the separate estimates are less likely to be statistically significant. The negative estimates for the oldest cohort are consistent with the relative supply effect of a larger cohort. The positive estimates for two younger cohorts and the older (50-59) cohort are not. Recall, though, that there is a potential positive bias in the estimates for cohorts for which migration is more related to labor market conditions – with in-migration to areas with stronger labor demand, boosting both relative cohort size and LFP. At the same time, the estimates for the older cohorts could be biased in the opposite direction from retirement-related endogenous migration.

Table 2.2: OLS and IV Regressions of Log Labor Force Participation Rate on Log Cohort Share, 1984-2016

Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Age	16-24	16-24	25-49	25-49	50-59	50-59	60-69	60-69
Both sexes								
ln(Cohort Size/Pop 16-69)	0.081**	-0.222	-0.036	-0.257**	0.032	0.295**	-0.074	0.385*
	[0.037]	[0.178]	[0.039]	[0.111]	[0.024]	[0.122]	[0.052]	[0.231]
IV: ln(Cohort Births/Births 16-69)								
Reduced-form coefficient (dep. var.=ln(LFPR))		-0.061		-0.086**		0.055**		0.103*
		[0.048]		[0.034]		[0.021]		[0.054]
1st-stage coefficient (dep. var.=ln(Cohort Size/Pop 16-69))		0.276***		0.336***		0.188***		0.267***
		[0.048]		[0.063]		[0.041]		[0.064]
Incremental R^2 (1st-stage)		0.046		0.015		0.005		0.011
1st-stage F-statistic		33.00		28.10		21.05		17.48
Durbin-Wu-Hausman test for endogeneity (p-value)		0.068		0.006		0.017		0.014
Men								
ln(Cohort Size/Pop 16-69)	0.014	-0.200	-0.031**	-0.114	0.012	0.241	-0.046	0.255
	[0.036]	[0.193]	[0.014]	[0.077]	[0.020]	[0.224]	[0.053]	[0.223]
IV: ln(Cohort Births/Births 16-69)								
Reduced-form coefficient (dep. var.=ln(LFPR))		-0.054		-0.041		0.031		0.071
		[0.052]		[0.027]		[0.026]		[0.064]
1st-stage coefficient (dep. var.=ln(Cohort Size/Pop 16-69))		0.272***		0.364***		0.130***		0.279***
		[0.051]		[0.064]		[0.047]		[0.058]
Incremental R^2 (1st-stage)		0.042		0.016		0.003		0.010
1st-stage F-statistic		28.51		32.59		7.75		23.57
Durbin-Wu-Hausman test for endogeneity (p-value)		0.242		0.238		0.241		0.152
Women								
ln(Cohort Size/Pop 16-69)	0.085*	-0.277	-0.059	-0.411**	0.071*	0.477***	-0.013	0.627*
	[0.043]	[0.200]	[0.067]	[0.187]	[0.042]	[0.183]	[0.044]	[0.338]
IV: ln(Cohort Births/Births 16-69)								
Reduced-form coefficient (dep. var.=ln(LFPR))		-0.079		-0.132**		0.099**		0.138***
		[0.054]		[0.057]		[0.039]		[0.050]
1st-stage coefficient (dep. var.=ln(Cohort Size/Pop 16-69))		0.284***		0.321***		0.207***		0.221***
		[0.054]		[0.068]		[0.041]		[0.075]
Incremental R^2 (1st-stage)		0.040		0.010		0.005		0.007
1st-stage F-statistic		27.63		22.31		25.97		8.66
Durbin-Wu-Hausman test for endogeneity (p-value)		0.047		0.011		0.019		0.005

Notes: Data source is described in notes to Figure 2.1, and specification details are described in notes to Table 2.1. The table reports estimates of equation (2.3.1). Regression weighted by average state population through the sample period. Standard errors are clustered by state. *, **, and *** denotes 0.1, 0.05, and 0.01 significance. The instrumental variable used for the cohort share is the total number of births within an age-cohort by state and year divided by the total number of births for 16-69 year-olds by state and year. We exclude certain years and states with missing birth data for the cohorts, as only a handful of states started reporting births in 1915. N = 1,326.

2.5.3 LFP: IV estimates

Table 2.2 reports the IV estimates of equation (2.3.1) for LFP, based on the level of the LFP rate for the cohort. Recall that constructing the IV causes us to lose the earliest years of the sample (plus some other earlier observations for some states). Thus, in Table 2.2 we first report OLS estimates for the same sample for which we can do the IV estimation. The OLS estimates are consistent with Table 2.1. For both sexes combined, we continue to find a positive and significant effect for the youngest cohort, a weaker positive effect for the 50-59 year-old cohort, and a negative although no longer significant effect for the oldest cohort. For 25-49 year-olds, however, the estimate is now negative. For men and women separately, the sign pattern is always the same, but there is some variation in which estimates are statistically significant.

The IV estimates tell a strikingly different story. For the two younger cohorts (16-24 and 25-49), the IV estimates point to a negative effect of relative cohort size on LFP, which is significant for the 25-49 year-old cohort, overall and for women. The estimated elasticities range from −0.114 to −0.411. These estimates are consistent with the standard relative supply hypothesis about the effect of relative cohort size. In every case (six estimations) the direction of change relative to the OLS estimates is consistent with positive bias induced by in-migration to stronger labor markets; i.e., the IV estimates become negative, or become more negative.

In contrast, for the two older cohorts (50-59 and 60-69) we find strong evidence of a large positive effect of relative cohort size, for both sexes combined, and for men and women separately (the estimates are significant for both sexes combined, and for women. The estimated elasticities range from 0.241 to 0.627. This evidence is inconsistent with the relative supply effect of a large cohort, and instead suggests that there are labor demand effects from large older cohorts that more than offset any supply effects. Like for the two younger cohorts, the IV estimates are quite different from the OLS estimates. However, for the older cohorts the direction of the change relative to the OLS estimates is in every case (again, six estimations) consistent with negative bias in the OLS estimates from endogenous migration related to retirement; the IV estimates become positive, or become more positive. Thus, the IV vs. OLS estimates are consistent with the kinds of biases we might expect – job-market related for younger cohorts, and retirement related for older cohorts.

Table 2.2 also presents additional information about the IV estimates. First, in each panel we report the reduced-form estimates – the effects on LFP of the relative cohort size variables defined based on births only. These always share the sign and significance of the IV estimates.[39] The reduced-form estimates tend to be roughly one-fourth of the magnitude of the IV estimates (this is also true for the analysis of wages below). As usual, the IV and reduced-form estimates answer different questions. For the purposes of asking what the behavioral response of LFP of older workers is to exogenous variation in relative cohort size, the IV estimate measures the parameter of interest. The reduced-form estimates capture solely the effects of variation in cohort size driven by the relative sizes of birth cohorts. Because there can be other sources of exogenous variation in cohort size (associated with, e.g., immigration and changes in industry structure – although not all this variation is exogenous), we regard the IV estimate as more relevant to asking, for example, what population aging implies for the likely LFP of older individuals.

Next, we report the 1st-stage coefficient estimates and F-statistics. The 1st-stage estimates are always positive and strongly statistically significant. The magnitudes are in the 0.13 to 0.36 range. The F-statistics are generally very large, ranging from around 8 to 33 (and only below 10 in two cases). Finally, we report p-values from the Durbin-Wu-Hausman endogeneity test. For the combined results and the results for women, these are below 0.05 in all cases but one, consistent with significant evidence of endogeneity bias.[40]

Finally, we also report the incremental R^2 from adding the instrument to the 1st-stage equations (including the fixed effects). For the older cohorts, it ranges from about 0.5% to 1.1%. (It is also high for the youngest cohort, consistent with a good deal of migration at young ages.) Consistent with the differences between the OLS and IV estimates, this suggests there is the potential for a good deal of endogenous variation, although of course much of the unexplained variation may not be associated with endogenous responses.

---

[39] Sometimes the significance level varies, but the same estimates are significant at the 10-percent or less.

[40] As noted above, we have estimated versions of the models in Table 2.2 (and Table 2.4 below) defining the size of the older cohorts relative to the younger cohorts (e.g., 16-49 year-olds). The results are very similar (available upon request).

Table 2.3: OLS Regressions of Log Average Hourly Wage on Log Cohort Share, 1979-2016

Sex	**Both sexes**				**Men**				**Women**			
Age	**16-24**	**25-49**	**50-59**	**60-69**	**16-24**	**25-49**	**50-59**	**60-69**	**16-24**	**25-49**	**50-59**	**60-69**
ln(Cohort Size/Pop 16-69)	-0.056*	0.226***	-0.066*	-0.029	-0.054*	0.181***	-0.040	-0.015	-0.009	0.209***	-0.050	-0.023
	[0.030]	[0.075]	[0.038]	[0.061]	[0.032]	[0.064]	[0.032]	[0.068]	[0.026]	[0.066]	[0.042]	[0.050]
Mean hourly wage (2016 dollars)	12.05	21.88	23.83	21.78	12.63	24.19	27.55	25.33	11.42	19.21	19.69	17.89
R^2	0.986	0.995	0.993	0.983	0.978	0.993	0.989	0.967	0.985	0.996	0.993	0.981
Incremental R^2	0.003	0.002	0.003	0.003	0.003	0.003	0.004	0.003	0.002	0.002	0.003	0.002

Notes: Data source and other details about the data are described in notes to Appendix Figure A5. See notes to Table 2.1. N = 1,938.

2.5.4 Wages: OLS Regression Estimates

We next turn to estimates of equation (2.3.1) for the effects of relative cohort size on wages. As reported in Table 2.3, for both sexes combined, we find a negative and significant effect of relative cohort size for 16-24 year-olds, with an elasticity of −0.056, consistent with the relative supply hypothesis. In contrast, for 25-49 year-olds the estimate is large and positive (an elasticity of 0.226). For the two older cohorts (50-59 and 60-69) the estimates are negative, fairly small, and statistically significant, at the 10-percent level, only for 50-59 year-olds (elasticity of −0.066). The magnitudes and the sign pattern of the estimates are the same for men and women separately.[41]

2.5.5 Wages: IV estimates

Table 2.4 reports the IV estimates for wages. As discussed above, there are less clear expectations regarding endogeneity bias in the estimated effects of relative cohort size on wages. First, while the younger and prime-aged cohorts may migrate to strong labor markets, the outward supply shift in these states may not do much to lower wages, and there can be offsetting effects from agglomeration externalities, and/or compensating differentials for congestion (e.g., Richardson, 1995). Second, for the older cohorts, as noted above, there is no clear prediction about bias from retirement-related migration.

In the IV estimates, which are of most interest, we again get a sharp message. For the two younger cohorts, there is evidence of a positive effect of relative cohort size, for the men and women combined, and for men. The elasticities range from 0.20 to 0.45 and are always statistically significant for both sexes combined and for men (in one case at the 10-percent level). For the two older cohorts, in contrast, the IV estimates always point to a negative effect – always significant for 50-59 year-olds (once at the 10-percent level), but not significant for 60-69 year-olds – consistent with the relative supply effect of a large cohort. The elasticities range from −0.16 to −0.56.

Like Table 2.2, Table 2.4 also reports diagnostic information about the IV estimates. The 1st-stage results are the same as for the LFP estimates, and hence are not reported again (see Table 2.2).

---

[41] The closest estimates in the existing literature are for the young cohort. Welch (1979) estimates elasticities of "entry" wages with respect to cohort size, for less-educated workers, in the −0.1 to −0.2 range. Macunovich (1999) reports an elasticity of about −0.07 with respect to size of birth cohort for young, less-educated workers. Morin's (2015) elasticities range from about −0.05 to −0.09 across age groups. However, estimation methods differ; Welch uses relative wages, while Morin and Macunovich used the level, and our estimates using a relative wage measure are less consistent with Welch (results available upon request). Our estimates also change when we instrument. Moreover, the results are not comparable for many reasons pertaining to differences in the analysis, included the much earlier data used in many of these other studies.

The reduced-form estimates always share the sign and significance of the IV estimates. And the p-values from the Durbin-Wu-Hausman endogeneity test indicates evidence of endogeneity bias in just over half the specifications. Although we did not have strong a priori expectations of endogeneity bias in wage estimates, the evidence suggests that sometimes there is such bias.

Table 2.4: OLS and IV Regressions of Log Average Hourly Wage on Log Cohort Share, 1984-2016

Estimator	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
Age	**16-24**	**16-24**	**25-49**	**25-49**	**50-59**	**50-59**	**60-69**	**60-69**
Both sexes								
ln(Cohort Size/Pop 16-69)	0.041	0.323**	0.089	0.450**	-0.012	-0.526***	0.019	-0.191
	[0.025]	[0.154]	[0.069]	[0.195]	[0.040]	[0.188]	[0.060]	[0.199]
IV: ln(Cohort Births/Births 16-69)								
Reduced-form coefficient (dep. Var.=ln(Average Hourly Wage))		0.089**		0.151**		-0.099***		-0.051
		[0.044]		[0.061]		[0.032]		[0.054]
Durbin-Wu-Hausman test for endogeneity (p-value)		0.062		0.011		0.003		0.229
Men								
ln(Cohort Size/Pop 16-69)	0.017	0.382*	0.048	0.397**	0.015	-0.559*	0.041	-0.155
	[0.030]	[0.204]	[0.059]	[0.178]	[0.030]	[0.314]	[0.076]	[0.230]
IV: ln(Cohort Births/Births 16-69)								
Reduced-form coefficient (dep. Var.=ln(Average Hourly Wage))		0.104*		0.144**		-0.073**		-0.043
		[0.054]		[0.061]		[0.033]		[0.065]
Durbin-Wu-Hausman test for endogeneity (p-value)		0.053		0.014		0.024		0.317
Women								
ln(Cohort Size/Pop 16-69)	0.042	0.198	0.056	0.308	0.020	-0.504***	0.004	-0.236
	[0.030]	[0.154]	[0.064]	[0.240]	[0.053]	[0.190]	[0.045]	[0.199]
IV: ln(Cohort Births/Births 16-69)								
Reduced-form coefficient (dep. Var.=ln(Average Hourly Wage))		0.056		0.099		-0.104***		-0.052
		[0.047]		[0.078]		[0.037]		[0.046]
Durbin-Wu-Hausman test for endogeneity (p-value)		0.349		0.216		0.002		0.236

Notes: Data source is described in notes to Appendix Figure A5. The 1st-stage estimates (and F-statistics) are the same as in Table 2.2. See notes to Table 2.2. N = 1,326.

2.5.6 Separate effects of size of older cohort relative to younger or prime-aged cohort

As it stands, then, the evidence on the estimated effects of a larger cohort on wages are cannot be fully reconciled with the estimated effects on LFP. Referring to the IV estimates, for the older cohorts, the LFP effects point to a positive demand shift towards older workers when the older cohort is larger, while the wage effects are negative (albeit insignificant for 60-69 year-olds); only the latter is consistent with the relative supply hypothesis. For the younger cohorts, in contrast, the LFP effects are most consistent with a negative relative supply effect, although the evidence is statistically significant only for 25-49 year-olds, while the wage effects are in the opposite direction. These contradictory findings are summarized in Table 2.6.

When we estimate the richer model (equation (2.3.2)) allowing for separate effects of the size of the older cohort relative to the two younger cohorts, we obtain a more coherent set of findings. These estimates are reported in Tables 2.5.A (for 50-59 year-olds) and 5B (for 60-69 year-olds). Here, we report the OLS and IV estimates for the consistent sample for which we can compute both. The IV estimations in Tables 2.5.A and 2.5.B are more demanding because there are now two endogenous variables. The 1st-stage F-statistics are often fairly large in both tables, although in some cases they do not exceed 10.[42] The IV results are qualitatively similar for all three samples – pooled, men only, and women only.

Looking first at the 50-59 year-old cohort, for LFP, the IV estimates indicate a weak negative effect or no effect of the size of the 50-59 year-old cohort relative to the youngest cohort (16-24), with elasticities ranging from −0.011 to −0.103. But for the size of the older cohort relative to the prime-aged cohort (25-49), the estimated effect is larger and positive in all three cases, and statistically significant in two of them; the elasticities range from 0.19 to 0.39. For wages, the effect of the size of the 50-59 year-old cohort relative to the youngest cohort is negative but not statistically significant, with elasticities ranging from −0.11 to −0.18. The estimated effect of the size of the older cohort relative to the prime-aged cohort is more strongly negative, and statistically significant in all cases, with elasticities ranging from −0.29 to −0.34.

Table 2.5.B presents similar estimates, for the oldest cohort of 60-69 year-olds. In the IV estimates, the sign pattern is identical to that for 50-59 year-olds. For LFP, the results are stronger statistically. There is a large, statistically significant negative effect of the size of the 60-69 year-old cohort relative to the 16-24 year-old cohort, with elasticities ranging from −0.44 to −0.48. And there is a large, positive and statistically significant effect of the size of the 60-69 year-old cohort relative to the 25-49 year-old cohort, with elasticities ranging from 0.46 to 0.77. For wages, only the estimated effect of cohort size relative to 16-24 year-olds for both sexes combined is statistically significant (at the 10-percent level); the elasticities range from −0.24 to −0.30.

[42] With multiple instruments the preferred diagnostic is Shea's Partial R^2, although these do not have a clear interpretation. And the Stock and Yogo (2005) critical values for minimum eigenvalues do not apply beyond homoskedastic i.i.d. errors.

Table 2.5.A: OLS and IV Regressions of Log Labor Force Participation Rate and Hourly Wages of 50-59 Year-Olds on Log Relative Cohort Sizes for 16-24 and 25-49 Year-Olds, 1984-2016

Estimator	**OLS**	**IV**	**OLS**	**IV**
Dependent Variable (Logs)	LFP	LFP	Hourly Wages	Hourly Wages
Both sexes				
ln(50-59 Cohort Size/16-24 Cohort Size)	-0.020	-0.061	0.019	-0.114
	[0.021]	[0.058]	[0.023]	[0.132]
ln(50-59 Cohort Size/25-49 Cohort Size)	0.038*	0.255***	-0.030	-0.318***
	[0.020]	[0.085]	[0.035]	[0.099]
1st-stage F-statistic				
(dep. var.=ln(50-59 Cohort Size/16-24 Cohort Size))		12.32		12.32
(dep. var.=ln(50-59 Cohort Size/25-49 Cohort Size))		11.61		11.61
Durbin-Wu-Hausman test for endogeneity (p-value)		0.001		0.002
Men				
ln(50-59 Cohort Size/16-24 Cohort Size)	-0.027*	-0.103	0.023	-0.179
	[0.014]	[0.088]	[0.018]	[0.210]
ln(50-59 Cohort Size/25-49 Cohort Size)	0.030	0.190	-0.014	-0.336**
	[0.019]	[0.120]	[0.029]	[0.157]
1st-stage F-statistic				
(dep. var.=ln(50-59 Cohort Size/16-24 Cohort Size))		5.83		5.83
(dep. var.=ln(50-59 Cohort Size/25-49 Cohort Size))		7.76		7.76
Durbin-Wu-Hausman test for endogeneity (p-value)		0.056		0.026
Women				
ln(Cohort Size/16-24 Cohort Size)	-0.019	-0.011	0.013	-0.092
	[0.029]	[0.100]	[0.030]	[0.136]
ln(Cohort Size/25-49 Cohort Size)	0.074**	0.392***	0.002	-0.293**
	[0.031]	[0.116]	[0.040]	[0.134]
1st-stage F-statistic				
(dep. var.=ln(50-59 Cohort Size/16-24 Cohort Size))		12.31		12.31
(dep. var.=ln(50-59 Cohort Size/25-49 Cohort Size))		10.71		10.71
Durbin-Wu-Hausman test for endogeneity (p-value)		0.002		0.002

Notes: Data source is described in notes to the figures, and specification details are described in notes to Table 2.1. The table reports estimates of equation (2.3.2). Regression weighted by average state population through the sample period. Standard errors are clustered by state. *, **, and *** denotes 0.1, 0.05, and 0.01 significance. The two instrumental variables used are the total number of births for 50-59 year-olds divided by the total number of births for 16-24 year-olds by state and year and the total number of births for 50-59 year-olds divided by the total number of births for the 25-49 year-olds by state and year. We exclude certain years and states with missing birth data for the cohorts, as only a handful of states started reporting births in 1915. N = 1,326. Note that we could use more observations in this table than in Table 2.5.B, for 60-69 year-olds, because the absence of early birth data is less of a constraint. However, we keep the samples the same in the two tables to make the estimates most comparable.

Table 2.5.B: OLS and IV Regressions of Log Labor Force Participation Rate and Hourly Wages of 60-69 Year-Olds on Log Relative Cohort Sizes for 16-24 and 25-49 Year-Olds, 1984-2016

Estimator	**OLS**	**IV**	**OLS**	**IV**

Dependent Variable (Logs)	LFP	LFP	Hourly Wages	Hourly Wages
Both sexes				
ln(60-69 Cohort Size/16-24 Cohort Size)	-0.120**	-0.459**	0.024	-0.239*
	[0.047]	[0.188]	[0.039]	[0.135]
ln(60-69 Cohort Size/25-49 Cohort Size)	0.077	0.597***	-0.010	-0.074
	[0.055]	[0.192]	[0.058]	[0.174]
1st-stage F-statistic				
(dep. var.=ln(60-69 Cohort Size/16-24 Cohort Size))		10.51		10.51
(dep. var.=ln(60-69 Cohort Size/25-49 Cohort Size))		9.72		9.72
Durbin-Wu-Hausman test for endogeneity (p-value)		0.006		0.083
Men				
ln(60-69 Cohort Size/16-24 Cohort Size)	-0.073	-0.480***	0.034	-0.299
	[0.047]	[0.163]	[0.043]	[0.183]
ln(60-69 Cohort Size/25-49 Cohort Size)	0.031	0.464**	0.007	-0.066
	[0.054]	[0.187]	[0.061]	[0.195]
1st-stage F-statistic				
(dep. var.=ln(60-69 Cohort Size/16-24 Cohort Size))		9.20		9.20
(dep. var.=ln(60-69 Cohort Size/25-49 Cohort Size))		14.20		14.20
Durbin-Wu-Hausman test for endogeneity (p-value)		0.013		0.101
Women				
ln(60-69 Cohort Size/16-24 Cohort Size)	-0.104**	-0.444	-0.045	-0.242
	[0.049]	[0.271]	[0.041]	[0.149]
ln(60-69 Cohort Size/25-49 Cohort Size)	0.124**	0.770***	0.052	-0.056
	[0.058]	[0.252]	[0.057]	[0.180]
1st-stage F-statistic				
(dep. var.=ln(60-69 Cohort Size/16-24 Cohort Size))		8.43		8.43
(dep. var.=ln(60-69 Cohort Size/25-49 Cohort Size))		4.85		4.85
Durbin-Wu-Hausman test for endogeneity (p-value)		0.002		0.129

Notes: Data source is described in notes to the figures, and specification details are described in notes to Table 2.1. The table reports estimates of equation (2.3.2). Regression weighted by average state population through the sample period. Standard errors are clustered by state. *, **, and *** denotes 0.1, 0.05, and 0.01 significance. The two instrumental variables used are the total number of births for 60-69 year-olds divided by the total number of births for 16-24 year-olds by state and year and the total number of births for 60-69 year-olds divided by the total number of births for the 25-49 year-olds by state and year. We exclude certain years and states with missing birth data for the cohorts, as only a handful of states started reporting births in 1915. N = 1,326.

Interestingly, then, when we look at the size of the two older cohorts (50-59 and 60-69) relative to the youngest cohort (16-24), the evidence is essentially fully consistent with the relative supply effect of a larger cohort – with negative effects on both LFP and wages. These results are again summarized, for men and women combined, in Table 2.6. In contrast, when we look at the size of the older cohorts relative to the prime-aged cohort (aged 25-49), there is relatively little statistical evidence for the relative supply effect of a larger older cohort. In particular, the LFP effect is positive for both older cohorts (in Tables 2.5.A and 2.5.B), and the wage effect is not significant and fairly close to zero 60-

69 year-olds in Table 2.5.B. The exception is the results for 50-59 year-olds in Table 2.5.A, where the wage effects are negative; but they are in the opposite direction of the LFP effects.

The striking finding here, in our view, is that when we break up the cohorts to which we compare the size of the older cohort, we get far less contradictory evidence of the effects of a larger older cohort. Table 2.6 helps illustrate this point in one table. As we see, when we simply looked at the relative size of the older cohorts (in Tables 2.2 and 2.4), using the simpler specification in equation (2.3.1), we found that the LFP effects of large older cohorts point to a positive demand shift towards older workers when the older cohort is larger, inconsistent with the relative supply hypothesis. The wage effects, in contrast, point to a negative relative supply effect..

In contrast, when we break up the cohorts to which we compare the size of the older cohort, using equation (2.3.2), all the evidence for the size of older cohorts relative to the youngest cohort fits the relative supply hypothesis. In contrast, almost none of the evidence for the size of the older cohorts relative to the 25-49 year-old cohort fits this hypothesis – and none of the evidence for LFP does. We highlight these results in Table 2.6.

How do we interpret the findings? The evidence of large negative effects on both LFP and wages for older workers aged 60-69, when their cohort is large relative to the youngest cohort, indicates that the oldest and the youngest workers are not very substitutable. Rather a large older cohort of 60-69 year-olds relative to 16-24 year-olds creates traditional, supply-side cohort crowding effects for older workers. This suggest that the effects are not driven by whether older workers taking post-retirement jobs move into jobs otherwise held by young people.

The results for the size of the 60-69 year-old cohort relative to the prime-aged cohort (25-49), however, are more consistent with a relative demand shift. There is a strong positive effect on LFP, suggesting that when the older cohort is large relative to the prime-aged cohort, demand for older workers is strong. When prime-aged workers are relatively scarce, firms may try to retain older workers. It is true that we do not find a corresponding positive wage effect for the older cohort; the estimates (in Table 2.5.B) are not significantly different from zero and they are small, although they are negative, rather than positive. While we cannot explain negative estimates via the demand side, if older workers' labor supply on the extensive margin is quite elastic, that could militate against finding a positive wage effect. And it is possible that the absence of wage effects or even negative effects, despite a positive demand shift, could arise from older workers entering into different kinds of employment relationships with their prior employers or new employers that are more flexible and pay less,[43] and in which they work fewer hours, or from negative selection on wages of who remains employed at older ages.

Table 2.6: Summary of Evidence on Relationship between Relative Cohort Size, and LFP Rates and Wages, IV estimates

Type of evidence	16-24	25-49	50-59	60-69
LFP				
Table 2.2, equation (2.3.1)				
Men and women	–	–**	+**	+*
Consistent with supply effect of relative cohort size?	Yes	Yes	No	No
Relative to two younger cohorts (Tables 2.5.A/B), equation (2.3.2)				
Men and women				
/16-24	…	…	–	–***
Consistent with supply effect of relative cohort size?	…	…	Yes	Yes
/25-49	…	…	+***	+***
Consistent with supply effect of relative cohort size?	…	…	No	No
Wages				
Table 2.4, equation (2.3.1)				
Men and women	+**	+**	–***	–
Consistent with supply effect of relative cohort size?	No	No	Yes	Yes
Relative to two younger cohorts (Tables 2.5.A/B), equation (2.3.2)				
/16-24				
Men and women	…	…	–	–*
Consistent with supply effect of relative cohort size?	…	…	Yes	Yes
/25-49				
Men and women	…	…	–***	–
Consistent with supply effect of relative cohort size?	…	…	Yes	No

Notes: See notes to Tables 2.2, 2.4, 2.5.A, and 2.5.B.

For 50-59 year-olds, the evidence for the effects of cohort size relative to the size of the youngest cohort (aged 16-24) is also no longer contradictory, as the estimated effects on LFP are negative estimates are negative (although not significant). The negative estimates are consistent with the conventional relative cohort size effect, like we found for 60-69 year-olds relative to 16-24 year-olds

[43] For example, Johnson et al. (2009, Table 1) reports that, among workers aged 51-55 in 1992, as of 2006, 14.2 percent remained at the same employer, 15.7 percent changed employer and stayed in the same occupation, and 26.9 percent changed employer and occupation. (The remainder were not employed.) And average wages are considerably lower on the new job (Table 17), which is typically less physically demanding, especially for those for those who changed occupations (Table 2).

(although the evidence was much stronger in this case). Only for the estimates for 50-59 year-olds relative to 25-49 year-olds does a contradiction remain, as we find positive estimates of the relative size of the older cohort on LFP, but negative and significant estimates on wages. Note, though, that the positive effects on LFP are the same as for 60-69 year-olds, although the magnitudes are smaller.

Thus, the disaggregation of the younger cohorts to a large extent resolves most of the contradictory evidence we found when lumping all "non-old" cohorts together. We find strong evidence, when compared to the size of younger cohorts, of traditional cohort crowding for workers aged 60-69. And when compared to the size of prime-aged cohorts, we found more evidence that large relative size of the oldest cohort is associated with a shift in demand towards older workers, although we cannot fully explain both the LFP and wage effects for the effects of the size of older relative to prime-aged cohorts in a simple demand and supply framework.

2.5.7 What do workers in older cohorts do when their younger cohorts are smaller?

The evidence from Tables 2.5.A and 2.5.B suggests that when the older cohorts of 50-59 or 60-69 year-olds are large relative to the 25-49 year-old cohort, LFP of the older cohorts is higher. This is consistent with an increase in demand for members of the older cohorts. Yet wages do not rise, which we speculated could be in part because the older workers induced to participate in the labor force when the younger prime-age cohort is smaller are entering into different kinds of employment relationships, possibly with lower pay. In this subsection, we present some evidence on this conjecture.

Panel A of Table 2.7 reports IV estimates of specifications similar to those in Table 2.5.A and 2.5.B, with the difference that we estimate models for the share of the labor force working part-time, or self-employed. If the LFP response among the older cohorts occurs via different kinds of employment relationships, then we might expect the shares part-time or self-employed to increase. Moreover, a self-employment response of this nature would be more likely to be for an unincorporated self-employed business – such as someone taking on a consulting role for a former employer. Hence, we also report specifications for the shares of the labor force in self-employment broken down by incorporation status. Aside from that, the approach is exactly as in Tables 2.5.A and 2.5.B, with the same first-stage, etc.

The estimates indicate that the margin of response for 50-59 year-olds to a smaller relative 25-49 year-old cohort is an increase in the share of the labor force working as self-employed. This effect is statistically significant (and is larger) only for the unincorporated self-employed, as hypothesized.[44] For 60-69 year-olds, Panel A of Table 2.7 indicates that the margin of response to a smaller 25-49 year-old cohort is an increase in the share of the labor force working part-time. Together, this evidence is consistent with older cohorts participating in the labor force at a higher rate, when their cohorts are large relative to 25-49 year-olds, in employment relationships that differ from common full-time, wage and salary arrangements. That is what we might expect given that the increase in LFP when old cohorts are relatively large come from those less attached to the labor force (and hence not participating when the relative size of older cohorts is not large).

Do these participation responses of older cohorts also explain the absence of positive wage effects (or even negative wage effects for 50-59 year-olds), in response to large cohorts of older workers relative to 25-49 year-olds? To explore this, Panel B of Table 2.7 simply reports regressions of our log average hourly wage measure on the shares of older workers in these alternative work arrangements. The evidence suggests that part-time work is associated with lower wages, although self-employment is not. Thus, these wage results provide a partial explanation for why the increase in LFP of older cohorts, when they are large relative to the 25-49 year-old cohort, is not accompanied by higher wages – as we would expect from a pure labor demand story. The explanation works for 60-69 year-olds – for whom the response occurs in part-time work – but not 50-59 year-olds.[45]

[44] There is also a smaller positive effect, significant at the 10-percent level, in response to a smaller 16-24 year-old cohort, although Table 2.5.A did not point to an increase in LFP for 50-59 year-olds when their cohort is large relative to 16-24 year-olds.

[45] We also estimated all of our models including two controls for the percentage of observations in either the leading (1946-1955) or trailing (1956-1964) edges of the Baby Boom (see https://www.census.gov/content/dam/Census/library/publications/2014/demo/p25-1141.pdf, viewed March 13, 2019). This had virtually no effect on the results (available upon request).

Table 2.7: IV Regressions of Part-time Work and Self-Employment of Older, and OLS Regressions of Log Average Hourly Wages on Log Part-Time and Self-Employed Workforce Shares 1984-2016

Older Cohort	50-59 Year-olds				60-69 Year-olds			
A. IV Regressions for Log Shares of Labor Force								
Dependent Variable, Log(·/Labor Force)	Part-time	Self-employed	Self-employed, incorporated	Self-employed, unincorporated	Part-time	Self-employed	Self-employed, incorporated	Self-employed, unincorporated
Both sexes								
ln(Older Cohort Size/16-24 Cohort Size)	0.306	0.554*	-1.408**	1.194**	0.062	0.037	0.125	-0.005
	[0.363]	[0.308]	[0.644]	[0.510]	[0.219]	[0.210]	[0.362]	[0.266]
ln(Older Cohort Size/25-49 Cohort Size)	0.371	0.511	-0.513	1.084**	0.614***	0.115	0.877	-0.030
	[0.251]	[0.323]	[0.465]	[0.528]	[0.232]	[0.291]	[0.676]	[0.329]
B. OLS Regressions for Log Average Hourly Wages								
Regression for part-time and for self-employed	-0.056***	0.004	…	…	-0.124***	-0.008	…	…
	[0.013]	[0.012]			[0.023]	[0.014]		
Regression for part-time, for self-employed incorporated, and self-employed unincorporated	-0.055***	…	0.007	-0.000	-0.123***	…	0.007	-0.016
	[0.012]		[0.007]	[0.010]	[0.023]		[0.009]	[0.009]

Notes: For Panel A, notes from Tables 2.5.A and 2.5.B apply. Each column is a separate specification. Part-time and self-employed come from different questions and are not mutually exclusive. Part-time and self-employment status are based on current employment only. In Panel B, each row reports results from two separate regressions for log average hourly wages, for 50-59 year-olds or 60-69 year-olds.

2.6 Conclusion

Our paper is motivated by the question of how employment (or labor force participation) and wages of older individuals are likely to change as the U.S. population ages, with a rising share of the population in the age ranges in which most people are retired. Couched in terms of the prior literature, this question concerns the effect of "cohort crowding" for older workers. When there is a relatively large cohort of older individuals, do we find that wages and labor force participation (LFP) are lower, because of the relative supply effect? Such evidence would be consistent with other research on younger workers. Or do we find different effects, perhaps because the age composition of the population affects the age composition of consumption and hence labor demand, or because a large relative cohort of older workers implies a small relative cohort of younger workers, which can itself affect demand for older workers?

We explore these effects of relative cohort size, taking account of the potential endogeneity of population structure owing to both work-related and retirement-related migration. We use as instrumental variables relative cohort size measures predicted by historical data on births in each state, by year. In general, we find evidence consistent with the kinds of biases we would expect from these two types of migration, and hence we emphasize the instrumental variables results.

When we study the effects of a large relative older cohort (aged 50-59 or 60-69) relative to the working-age population as a whole, we find contradictory evidence. For LFP, we find evidence that is inconsistent with the relative supply or cohort crowding hypothesis, and which instead suggests an increase in demand for older workers when the older cohort is relatively large (with higher LFP). But we find negative wage effects, consistent with the relative supply hypothesis.

However, when we look at the size of the older cohorts relative to a young cohort (aged 16-24) and a cohort spanning the prime/middle range of ages (25-49) we find a more coherent set of results. When the older cohort is large relative to the younger cohort, the evidence is much more consistent with the relative supply hypothesis, with a larger relative older cohort reducing LFP and wages. But when the older cohorts are large relative to the cohort of 25-49 year-olds, LFP of older workers is higher, and it is less clear that wages are affected.

These results for the size of older cohorts relative to prime-aged cohorts are more consistent with a relative demand shift. When prime-aged workers are scarce relative to older workers, firms may try to retain or hire older workers. Older workers' extensive margin labor supply elasticity may be quite high. Moreover, older workers often enter into different jobs or employment relationships with more flexible, lower-paying work. There is some evidence from data on part-time work and self-employment that the increase in older workers' LFP when their cohort is large relative to the 25-49 year-old cohort comes via self-employment or part-time work. Moreover, for 60-69 year-olds this may help explain why average wages do not rise despite the increase in LFP; for 50-59 year-olds, in contrast, there remains more of a contradiction between higher LFP but lower wages when their cohort is large relative to 25-49 year-olds.

Together, the results suggest that cohort size may have important implications for the LFP (and wages) of older workers. However, our evidence suggests that we need a more nuanced view than simply whether the older cohort is large relative to the population: The cohort they are large relative to matters. Our evidence also suggests the value of additional work to understand the behavior underlying our findings, both to better understand the labor market decisions of older workers, and to assess the validity of the interpretation of the results we find in this paper.

Nonetheless, as it stands, the pattern of projected population aging is most consistent with rising shares of 50-59 year-olds and 60-69 year-olds relative to the broad group in their 20s, 30s, and 40s, rather than an increase relative to particularly small young cohort.[46] As such, our results suggest that population aging is likely to be accompanied by rising labor force participation and hence employment of older individuals.

That said, there are some potential limitations to our analysis, which remain for future research. We motivated our analysis by asking what changes in the age structure of the population imply – first and foremost – for labor force participation of older individuals. Our estimates based on the IV we use are informative for this interpretation, which can also be thought of as extrapolating from our estimates to project the likely effects of population aging in the aggregate data, which can be viewed as exogenous. Our estimates do not, however, disentangle or decompose the labor supply and demand responses of those at different ages to population aging. Moreover, there is potentially a rich set of life-cycle responses in which large cohorts who are now older engaged in when they were younger, such as increased educational investments owing to lower wages (e.g., Berger, 1984), leading to higher employment at older ages to recoup the earlier educational investments. That is, there is potentially rich and interesting "black box" of behavioral responses to changes in age structure that we do not explore, although doing so goes well beyond the purpose of this paper.

[46] See https://census.gov/data/tables/2017/demo/popproj/2017-summary-tables.html (viewed March 13, 2019).

3. Does Rapid Transit and Light Rail Infrastructure Improve Labor Market Outcomes?

3.1 Introduction

3.1.1 Motivation

I evaluate the effects of expanding the Los Angeles Metro rapid transit network of light rails, subways, and bus rapid transit (BRT)[47] on labor force participation and employment in the neighborhoods these stations are placed in. Los Angeles has been consistently expanding its Metro network since the initial construction of a light rail line in 1990, following the absence of the last consumer streetcar in 1963, to the completion of 103 Metro stations by 2016. This paper is primarily motivated by the fact that public transit has often been proposed as a solution to "spatial mismatch", historically defined by economists as a mismatch between where low-income households reside and suitable job opportunities. Spatial mismatch traditionally focuses on the effects of residential segregation in the urban core for low-skilled and minority workers and the suburbanization of job opportunities far away from the urban core (Kain, 1968; Kain, 1992; Gobillon et al., 2007). While researchers and policymakers have often proposed public transportation improvements to address spatial mismatch, it remains unclear whether public transit improvements have a significant effect on the labor market outcomes.

The papers that address the link between public transit and employment are limited and generally focus on accessibility of transit in terms of cost (Phillips, 2014; Rodriguez et al 2015), racial hiring decisions of employers as a result of increased transit (Holzer et al. 2003) or have weak identification strategies that do not fully address endogenous placement of routes (Sanchez, 1999; Sari 2015; Heilmann 2018). This paper provides several contributions. First, it expands on a sparse literature tying public transit infrastructure to employment outcomes and addresses endogenous placement of routes that has not been addressed in previous studies. Additionally, from a public policy perspective, the infrastructure expansion was costly and voted directly by referendum to increase gas taxes in Los Angeles County. By evaluating the effects on the labor market, this paper addresses whether there were tangible benefits to funding Metro expansions. Finally, it addresses the spatial

[47] The decision to include the Orange line, a bus-rapid transit system, is because of its dedicated right-of-way and an on-time performance of 94%, close to the 99% of the train system (Mendelson 2015)

mismatch component by observing whether transit accessibility can increase employment and reduce the disparities in labor market outcomes among different subgroups more vulnerable to this mismatch.

Using Los Angeles as a case study also addresses the spatial mismatch hypothesis better than other cities, because the high concentration of minority groups and large variations in income across neighborhoods allows the evaluation of the differential employment effects across racial and low-income groups. Additionally, the Los Angeles County Metropolitan Transportation Authority (MTA) is attempting to finish construction of a series of expansions by the 2028 Summer Olympics. The effects of long-run infrastructure improvements on labor market conditions resulting from sporting events, such as the expansion of public transit, has not been fully explored in the literature, which tends to focus on the short-run effects of sporting events on employment.[48] Much of the literature on public transit infrastructure improvements, focuses on how Metro stations have a gentrification effect by increasing housing prices. (Kahn 2007; Bowes and Ihlanfeldt 2001; Zheng and Kahn 2013; Hu 2017; Dziauddin 2019; Zhou et al. 2020). However, there's little work on how the impact of public transit infrastructure can increased commercial activity and sprout new businesses in the areas, thereby improving the employment opportunities for their residents. While one may think of public transit increasing employment by connecting residents to job opportunities that were otherwise inaccessible due to commuting costs, it can also increase job density in the neighborhood and reduce spatial mismatch.

To briefly summarize the method and results of the paper, I first evaluate the effects of proximity to a station on public transit usage, labor force participation, and employment using a standard panel data model with tract and time fixed effects. However, route placement may be endogenous to changes in existing demand for public transportation and neighborhood composition. Thus, I calculate the distance of the centroid of each tract to a hypothetical Metro route as an instrument to predict new station locations. The hypothetical Metro route is modelled from a least-cost route that minimizes elevational slope changes to connect the origin and destinations of Metro stations. Overall, I find that proximity to Metro stations increases labor force participation and employment, reducing spatial mismatch. The increased employment is likely driven by increased job densities in neighborhoods close to Metro stations rather than allowing residents to connect to distant job opportunities.

---

[48] A large consensus seems to be that the positive employment effects are minimal, at best, and often fall short of initial projections or statistics published by governmental sources (Baade and Matheson, 2002; Hotchkiss et al., 2003; Baumann et al.2012).

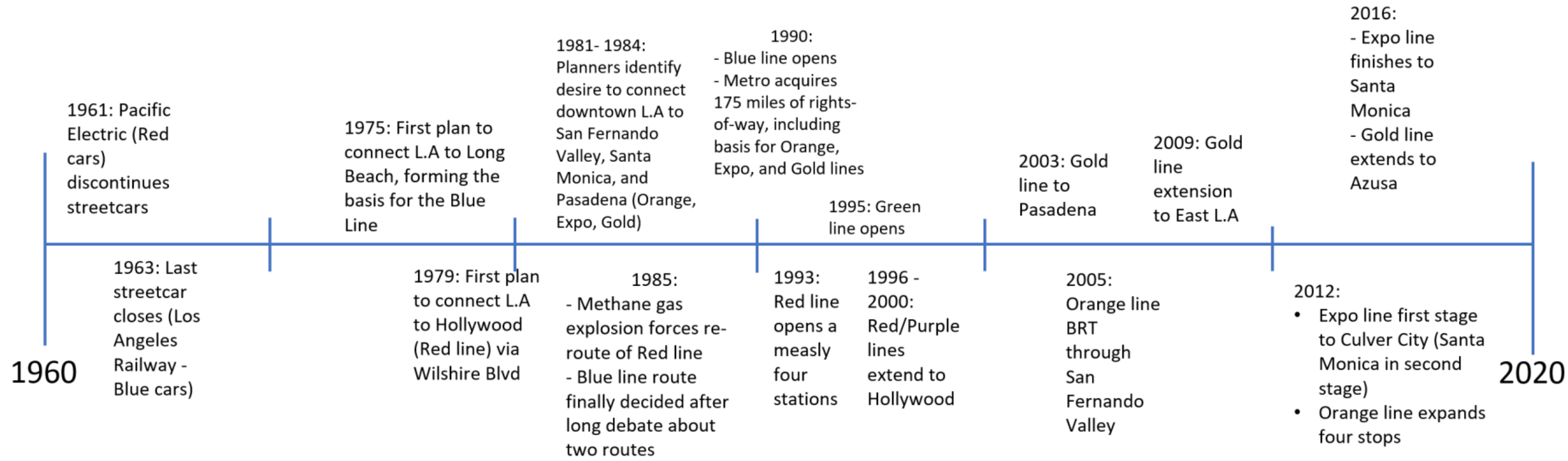
3.1.2 Brief History of the Los Angeles Metro

In the early 1900s, Los Angeles had an extensive urban electric streetcar network primarily owned by two companies, Pacific Electric and the Los Angeles Railway, colloquially known as the red cars and yellow cars. However, due to the proliferation of automobiles leading to congestion, the unprofitable fare restrictions and maintenance, and allegations of inference from the automobile industry to switch to buses, all streetcar systems closed by 1963 (Marshall 2016). Figure 3.1 displays a timeline of relevant events

In 1975, there were initial talks of constructing a modern rapid transit system in Los Angeles. For the next few decades, planners identified areas that they wished to connect by transit, but a series of political, legal, and logistical hurdles significantly delayed construction. The first route, the Blue line, which had been proposed since 1975, finally became operational in 1990. Soon after, the Red/Purple line was constructed in two stages, in 1993 and 1996 due to some issues with route placement that will be discussed later, and the Green line was constructed in 1995. However, despite the acquisition of a wide network of 175 miles of former Pacific Electric rights-of-way in 1990, development soon fizzled out due to budgetary issues, the economic downtown in the early 1990s, and the decreased public standing of the Metropolitan Transit Authority (MTA). More importantly, residents in the areas initially mobilized heavy resistance against transit routes due to concerns about neighborhood safety.

However, in the early 2000s, there was a shift in public opinion back towards developing rapid transit due to increased traffic, an increase in federal and state grants, and the increase in politicians favorable to transit (Elkin 2014). This led to the development of the Gold line in 2003 that connected Union Station to Pasadena, the Orange bus rapid transit system (BRT) in 2005 from Hollywood through the San Fernando Valley, the Gold line light rail expansion from Union Station to East Los Angeles in 2009, an expansion to the Orange line BRT in 2012, the first leg of the Expo light rail construction in 2012 that connects downtown to Culver City, the expansion of the Expo line in 2016 that connects to Santa Monica, and the expansion of the Gold line from Pasadena to Azusa in 2016. Additionally, in 2008, the passage of a proposition, Measure R, secured $40 billion dollars for MTA over thirty years, with 35% dedicated to transit capital projects. This measure helped guarantee the future for its current and planned projects through 2026. Figure 3.2 shows all routes in 2000, 2016, and the planned, future routes through 2026 (L.A. Metro).

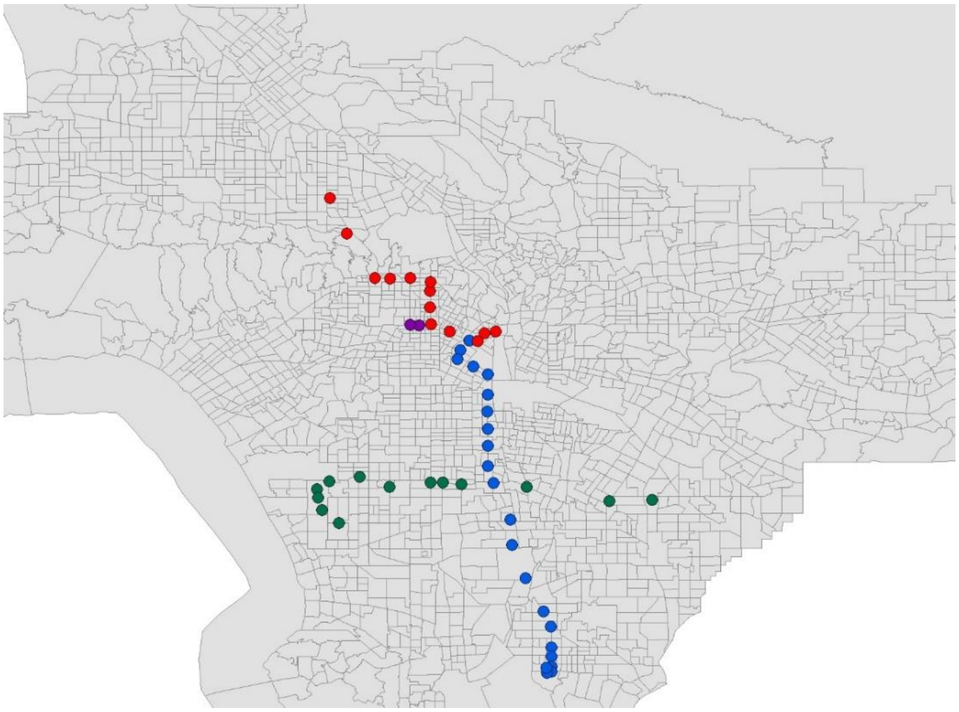
Figure 3.1: Timeline of the Los Angeles Metro:



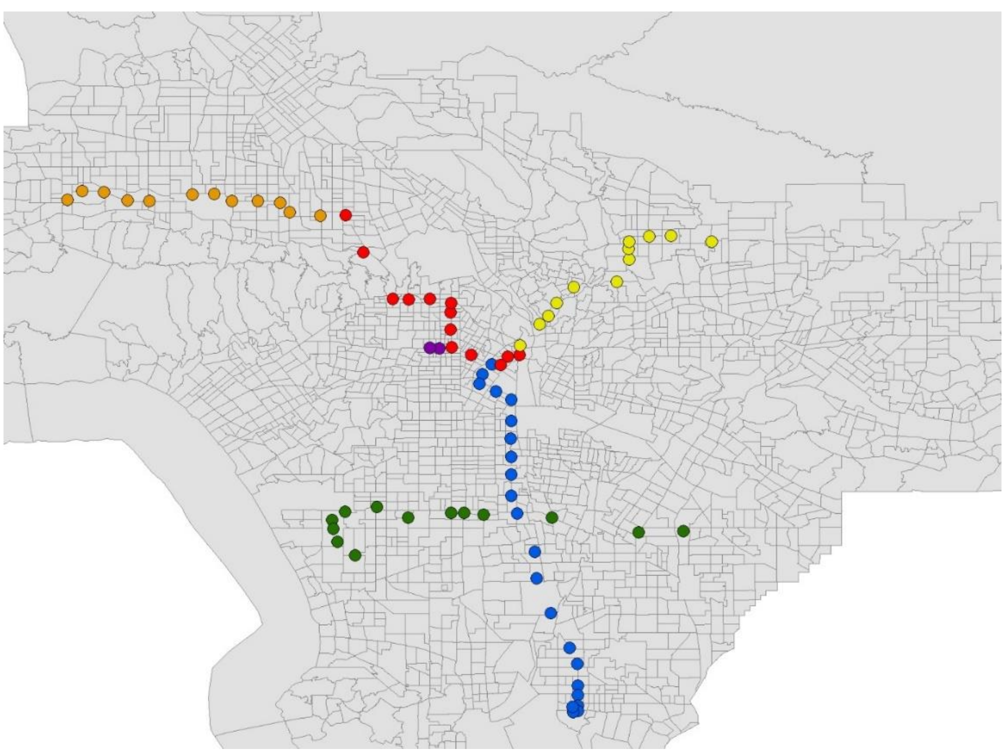
Sources: Elkin (2014) and L.A. Metro

Figure 3.2: Los Angeles Metro Stations in 2000, 2005, 2012, 2016, and Planned 2026 Routes

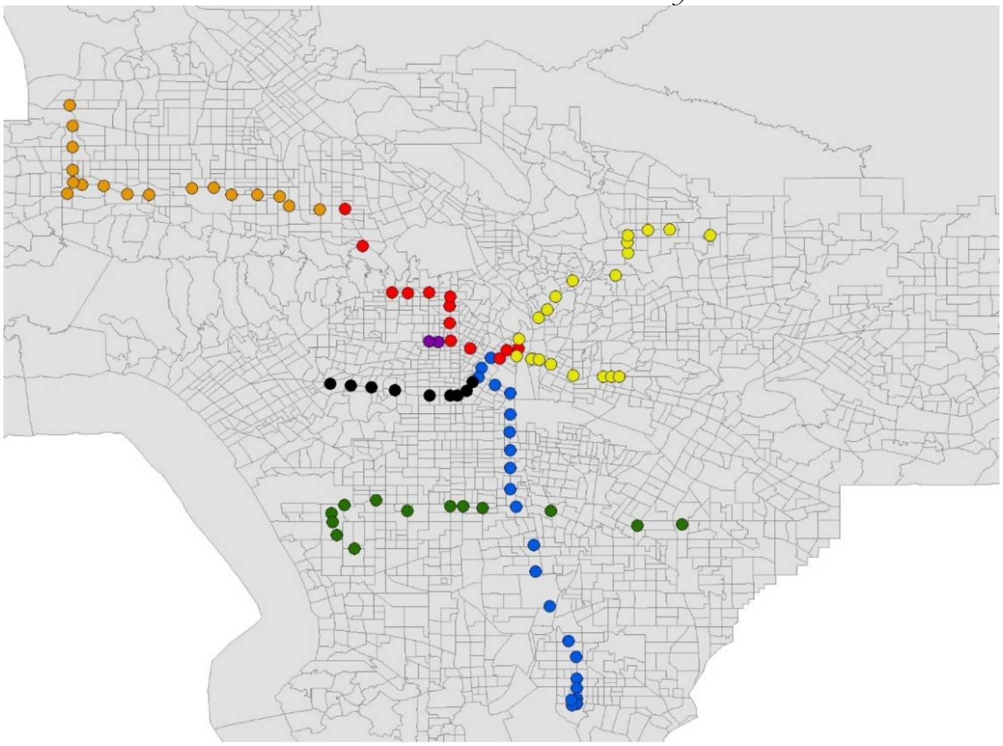
A. L.A Metro Stations in L.A. County in 2000



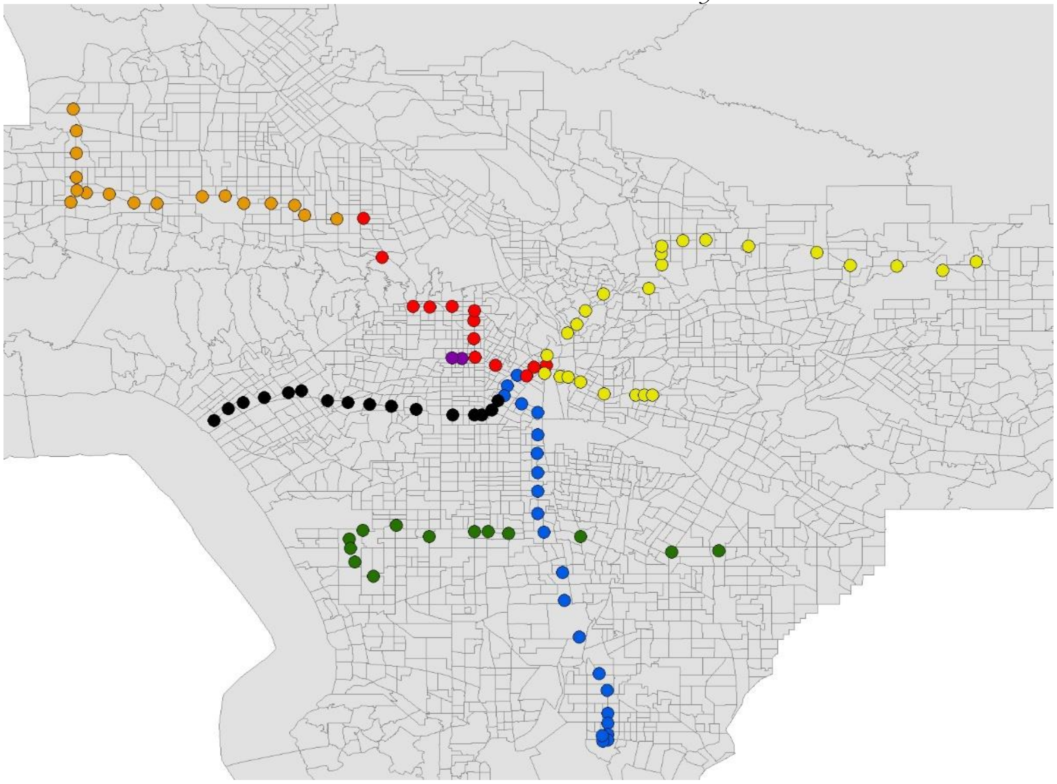
B. L.A Metro Stations in L.A. County in 2005



C. L.A Metro Stations in L.A. County in 2012

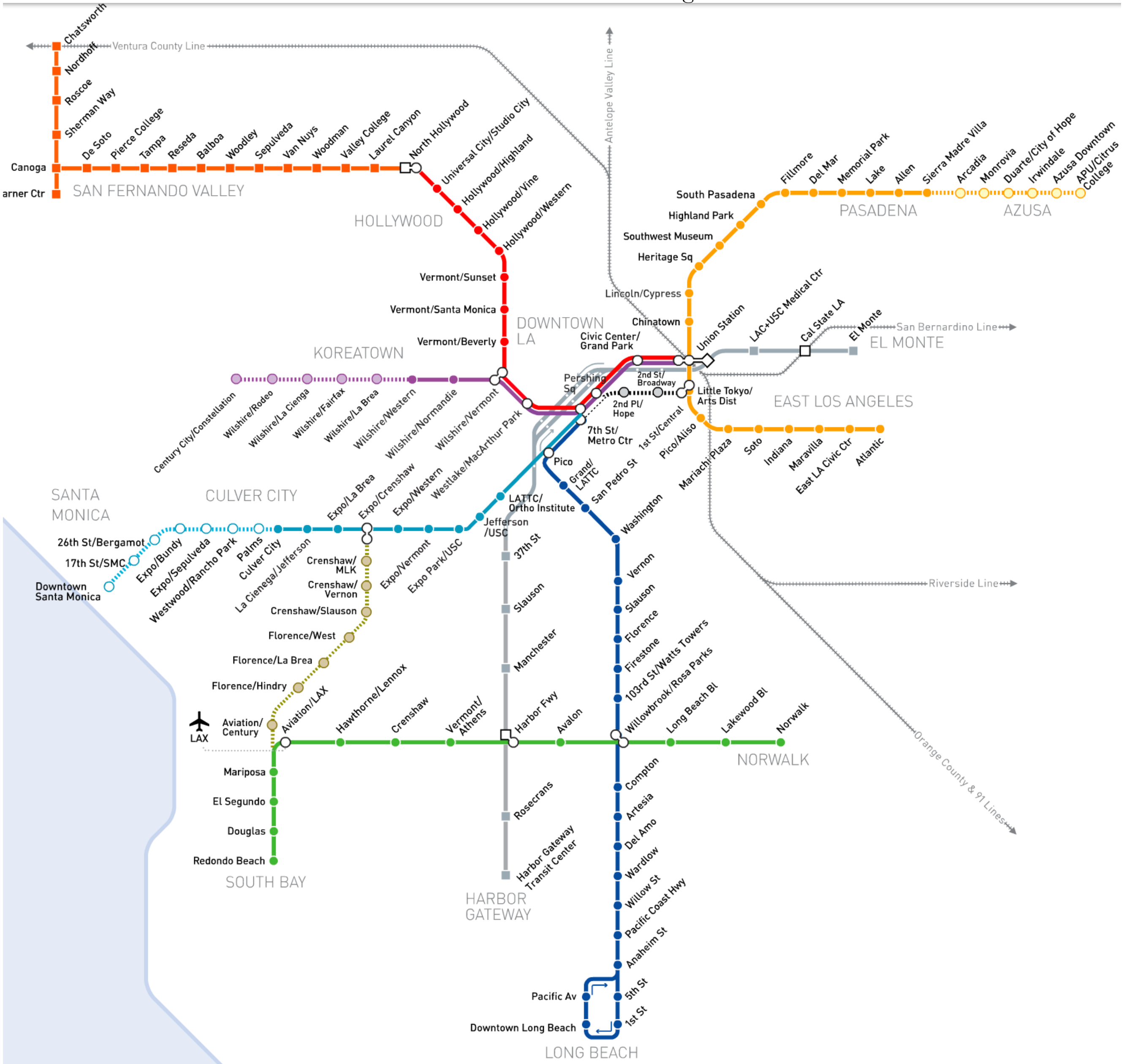


D. L.A Metro Stations in L.A. County in 2016



Source: Los Angeles Metropolitan Transit Authority overlaid on Los Angeles County shape file. The black routes are the Expo line and the yellow is the Gold line – all other routes are colored according to their name.

E. L.A. Metro Station Planned Routes through 2026



Source: Los Angeles Metropolitan Transit Authority.

3.2 Literature Review

3.2.1 Spatial Mismatch

I divide the literature review into two sections, one introducing the concept of spatial mismatch and the other on the current literature regarding public transit and employment. Before spatial mismatch was first introduced, Mills described a model of urban residential choice that assumes individuals choose optimal residential locations based on employment in a fixed city center and face tradeoffs between housing costs and commute times based on their choice of residency (Mills 1972). However, due to the decentralization of employment and individual preferences for amenities, there may be a mixture of high and low-income individuals who reside in the central city or the suburban areas (White 1988) and residents living in the suburbs generally have higher incomes to compensate for longer travel times (Madden 1985).

However, many people face barriers to obtaining their preferred residential locations. Kain (1968) explained the high unemployment rate for Blacks as stemming from housing segregation and the suburbanization of job opportunities. Kain's paper focused on Chicago and Detroit and how low-skilled jobs moved from the urban core of cities to the suburbs, while Black workers were still constrained to the urban core due to housing market discrimination. This idea was later dubbed "spatial mismatch", a disconnect between where low-income residents reside and suitable job opportunities, and a large subsequent literature stressed the importance of proximity to workplace or accessibility of jobs on employment.[49]

Gobillion et al. (2007) suggested seven mechanisms why greater distances to job opportunities could be harmful. On the worker's side, four possible explanations are that workers may refuse a job that involve long and costly commutes, workers' job search efficiencies may decrease with distance to jobs, workers residing far away from jobs may not search intensively, and workers may incur high search costs that restrict their spatial search horizon to nearby neighborhoods. The first mechanism can be explained by higher reservation wages for costly commutes, resulting in lower employment for individuals far away from job opportunities (Brueckner and Zenou 2003; Coulson et al. 2001). Additionally, barriers to search exist from the reduced availability of job information (Ihlanfeldt and Sjoquist 1990, Ihlanfeldt and Sjoquist 1991) and workers may be deterred from high search costs and restrict their search to nearby neighborhoods (Stoll 2005, Holzer and Reaser 2000). Under these labor supply side explanations, improvements in reliable public transit methods should decrease commuting costs and increase search intensity, resulting in a positive effect on employment.

On the labor demand side, Gobillion et al. (2007) also suggested three firm-related mechanisms for spatial mismatch. Employers may discriminate against residentially segregated workers because of stigma or prejudice associated with their residential location (statistical discrimination), employers may refuse to hire or prefer to pay lower wages to distant workers because of the fear that long commutes will make a worker less productive, and suburban employers may think

[49] For a detailed literature review on spatial mismatch papers, see Gobillion et al. (2007) and Holzer (1991)

their White local customers are unwilling to have contact with minority workers (customer discrimination). There is some evidence that firm-based discrimination may occur. Hellerstein et al. (2008) suggested the problem is less of a spatial mismatch but rather a "racial mismatch", where there is a lack of jobs where minority groups are hired into. Phillips (2014) runs a correspondence study and finds employers discriminate against employees who live further from job locations. Under the first and third scenarios that have a strong racial discrimination component, improvements in public transit will likely have a null effect on employment, as transit accessibility will not necessarily address these problems.[50] The second scenario regarding the employer's fear of decreased worker productivity is a bit unclear; public transportation can reduce commuting costs which should increase the perception of worker productivity.

3.2.2 Public Transportation as a Solution to Spatial Mismatch

Spatial mismatch papers often recommend improvements in public transit infrastructure and accessibility as a potential solution to addressing the issue (Kain 1968). However, despite this insistence on public transit infrastructure improvements as a viable solution, the evidence in the literature addressing the connection between public transit infrastructure and employment is somewhat limited in their identification strategies.[51] For example, while a few early studies find a positive association between public transit and employment at the intensive margin (Sanchez 1999; Kawabata 2003), none of these studies sufficiently address endogeneity issues, relying on regression control and cross-sectional data. Holzer et al. (2003) find that firms located near transit stations in San Francisco hired more Latino workers, but this type of analysis focuses on the decision-making process of firms rather than the benefits to residents of an expanded transit network.

More recently, there have been papers that use difference-in-differences or two-way fixed effects models to address the link between public transit and employment (Sari 2015; Heilmann 2018).[52] Sari (2015) finds a new tramway in Bordeaux, France decreased unemployment rates in neighborhoods located close to the tramway station more relative to other neighborhoods. For an

[50] There may be an exception if public transit improvements increase job density for jobs that minorities can be hired into, reducing the racial mismatch component.

[51] There is a growing literature focusing on the effects of reducing transit accessibility in terms of cost on employment outcomes using randomized controlled trials (see Phillips 2014; Rodriguez et al 2015). My study focuses on transit accessibility in terms of the availability of infrastructure rather than the cost of using public transit.

[52] Aslund et al (2017) find a zero impact of commuter train access in Sweden on earnings and employment. However, intercity commuter trains are fundamentally different from rapid intracity transportation systems, and I do not explore similar systems such as the Metrolink in this paper, mainly due to data limitations.

example from the United States, Heilmann (2018) find positive effects on income from the construction of Dallas's urban rail network. However, the methodology could be problematic if station placement is correlated with changes in existing demand for public transit infrastructure. In Heilmann (2018) for the case of Dallas's train, the initial plan was to construct over 160 miles of rail, which was later curtailed to 93 miles of rail. Thus, it is possible that Dallas planners may have chosen which plans to curtail depending on the changes in economic conditions, ridership, or transportation demands in specific neighborhoods. In my study, I address any potential biases from Metro line placement decisions with instrumental variables, an approach that has not been used in these previous studies. Severan (2019) studied welfare effects resulting from the construction of the LA Metro system. However, there are a few notable differences. First, Severan focuses on different time periods and uses 1990 and 2000 census data to evaluate the first few routes constructed in the 1990s. Secondly, Severan builds a spatial general equilibrium model that accounts for commuting, housing markets, and labor markets to determine the overall welfare effect. While both Severan and this paper study the same setting, ultimately, the papers take different approaches in both their estimation strategy and outcome of interest. My paper's goal is to address the spatial mismatch question by estimating the average employment and earnings effect for residents and whether this is attributed to increased job density nearby.[53]

Finally, while transit infrastructure improvements may address spatial mismatch by connecting residents to farther-away jobs that were otherwise inaccessible, thereby increasing overall employment, residents may also increase their employment from increased job opportunities in areas near transit stations. However, the literature on that is unclear. Credit (2017) finds growth in new businesses in the knowledge, service, and retail sectors after station openings from the Phoenix light rail system, although the effect tapered off over time. Schuetz (2015) do not find positive effects of Metro stations on retail activity in four California MSAs, including Los Angeles from 1992-2009. The follow-up to this paper suggests that the retail growth may be lagged 5-10 years (Schuetz et al. 2018). However, there is a larger and more consistent literature on the gentrification effects of public transportation improvements. Overall, these papers find that proximity to transit stops increases property values or housing in the area (Kahn 2007; Bowes and Ihlanfeldt 2001; Hu 2017; Dziauddin 2019; Zhou et al. 2020).

---

[53] There are some subtle similarities in minor choices, such as using walking distance buffers (although a different distance threshold was chosen) and using historical routes as control tracts. These were determined in this paper without the author's knowledge of Severan's working paper.

3.3 Data

I use publicly available data from the American Community Survey (ACS). These summary files contain averages for labor market outcomes within a tract, such as labor force participation, employment to population ratio, median income earned in the past 12 months, and hours worked. The ACS also has data on the number of workers who use public transit, broken down by bus and subway/rail, which I divide by the resident population. Additionally, I retrieve demographic data such as race, gender, education, migration, and vehicle ownership as controls. The main limiting factor with the publicly available ACS data is that it contains information at the census tract level for aggregated 5-year samples only, for confidentiality concerns. For example, the employment data reported in 2009 contain the average employment by tract from 2005-2009. This is limiting in that it only allows for one pre-intervention and one post-intervention period. Therefore, I evaluate the 2009 and 2012 expansions by using the 2005-2009 as the pre-period and 2012-2016 as the post period.[54] The 5-year file for 2005-2009 is converted from 2000 census tracts to 2010 census tracts using the Longitudinal Tract Database (LTDB).[55]

Ideally, I could identify the tract of work specifically from the ACS data to determine whether there are increased job densities. However, this has been restricted to confidential data. Therefore, I use Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) data to measure the effects of proximity to Metro stations on job density. LODES provide the total number of jobs in each origin-destination pair and distinguishes the type of job by income (low, medium, or high).[56] Low income jobs are defined as jobs with monthly earnings of $1250 or less, medium-income jobs are classified as monthly earnings of $1251 to $3333, and high-income jobs are defined as monthly earnings greater than $3333. These definitions are from the LODES data directly and not by construction. One benefit of LODES is that unlike the public ACS data, it is disaggregated down to the block level by year, allowing for a long panel of yearly observations from 2002-2016, which enables the pooling of all Metro expansions. However, the main results I present in this paper

---

[54] I also tested evaluating each Metro expansion in 2009 and 2012 separately (i.e. using 2010-2014 as the post period for the 2009 expansion and 2007-2011 as the pre-period for the 2012 expansions) but along with issues regarding overlapping Metro expansions, for the most part, the results were not too different.

[55] LTDB uses an interpolation based on both geographical area and population weights. For a discussion of the benefits of LTDB over the Neighborhood Change Database (NCDB), see Logan et al. (2014). Lee and Lin (2018) also prefer LTDB over NCDB for converting 2000 tracts into 2010 tracts.

[56] There's also data on age group (less than 29, 29-55, and over 55) and industry, but they did not generate any interesting results

are aggregated to a two-period model to allow for the same control variables from the ACS and an apples-to-apples comparisons with the specification using the ACS.[57]

I retrieve information about the location and timing of Metro stations from the LA Metro website. Addresses are geocoded into ArcMap and overlaid on top of census tract shape files. ArcMap allows a variety of spatial analysis tools, such as the calculation of distances from a point or a line to the centroid of the tract and the calculation of least-cost paths, which will be explained in the econometric specifications.

3.4 Estimation

3.4.1 Difference-in-differences (DiD)

The model in equation (3.4.1) represents a simple two-way fixed effects regression with tract and time dummy variables that allows for the pooling of multiple Metro expansions and variation in the data across many time periods.

$$Y_{ct} = \alpha + \beta_1 * Metro_{ct} + \beta_2 * X_{ct} + \Upsilon_c + \lambda_t + \varepsilon_{ct} \qquad (3.4.1)$$

In equation (3.4.1), $Metro_{ct}$ indexes whether a tract c contains a Metro station at time t within walking distance. Walking distance has traditionally been defined in the urban planning literature as 0.4 kilometers from transit stops although some studies suggest that rail transit may be higher than 0.8 kilometers.[58] For the purpose of this study on rapid transit systems, I use a 0.8 km buffer around the Metro station.[59]

Y_{ct} indexes the outcomes of interest. From the ACS, I measure the effect of having a Metro station within walking distance on average labor force participation, employment, public transit usage, and median income earned in past twelve months within a tract. From the LODES, I measure the effect of having a Metro station within walking distance on job density, the number of jobs in a tract. It is possible that a Metro stop may lead to growth in available jobs in the tract through increased

[57] I also tested the specifications using a multi-period model. However, the results were not substantively different and by doing so, I cannot use the control variables in the ACS for an analogous comparison. Additionally, the logic for the instrument is a bit different, as it is no longer interacted with post but use a new method to create variation across all time periods. For these reasons, I omit the multi-year model in this paper.

[58] For a detailed literature review on the usage of different walking distance buffers to transit, see El-Geneidy et al. (2014)

[59] Using a smaller 0.4 km buffer yielded similar results. Results available upon request.

commercial activity or new businesses.[60] The dependent variable is expressed in terms of actual job levels and a ratio of jobs per population (population taken from the ACS).[61]

The controls in X_{ct} include shares of Black, female, high school degree, some college, bachelor's or higher, out-of-state migration in the past year, out-of-county migration in the past year, no vehicles, and only one vehicle. Including shares of out-of-state and out-of-county migration within the past year can address a potential endogeneity concern that people already participating in the labor force may decide to migrate to tracts near the stations after they are constructed although this would be better if within county migration were available. Confidential ACS data has more detailed migration data on migration by tract within the past year, and excluding these individuals can serve as a strong robustness check to partially rule out neighborhood sorting mechanisms. λ_t indicates a vector of dummy variables for time and Υ_c indicates a vector of dummy variables for census tracts. Because I use sample estimates of tract-level averages to construct the data, I weight the regression by the average population of the studied group (i.e total population (16+) when analyzing labor force participation for the total population, Black population when analyzing labor force participation for Blacks, etc.).[62] While the mean population size of a tract is 3346, there are a couple of tracts with very small population sizes (see Figure 3.3). Therefore, I removed all tracts under 100 population. One potential shortcoming with equation (3.4.1) as specified above is that is assumes constant treatment effects, whereas we may think that the treatment effects depend on the size of the network at a given time. A more general specification would allow for an interaction between $Metro_{ct}$ and the size of the network.

However, given that the ACS data is only given in 5-year intervals, I am restricted to two-periods rather than the multiyear model in equation (3.4.1). I also account for network effects by estimating a more simplified version of the general model that considers two types of stations in two periods – existing stations in operation in the pre-period and new stations that began operation in the post period.[63] Therefore, I modify equation (3.4.1) to the following:

---

[60] Because LODES gives a place of residency and work, I also evaluated changes in commute distances and the change in the percentage of workers who work in tracts with Metro stations but did not find the results that meaningful due to the inconsistencies with the ACS and the inability to isolate out specific groups. The ACS microdata can provide a more detailed analysis on these topics.
[61] I also tested using the inverse arcsine of jobs to calculate a percentage change, but I did not think it was prudent to place greater emphasis on for large percentage changes to jobs in smaller tracts.
[62] I test out each specification without the population weights as well. In the end, the resulting estimates do not seem to differ much. Results for unweighted regressions are available upon request.
[63] In two periods, the $post_t$ interaction fully accounts for the increase in the size of the network.

$$Y_{ct} = \alpha + \beta_1 * New_c * post_t + \beta_2 * Existing_c * post_t + \beta_3 X_{ct} + \Upsilon_c + post_t + \varepsilon_{ct} \tag{3.4.2}$$

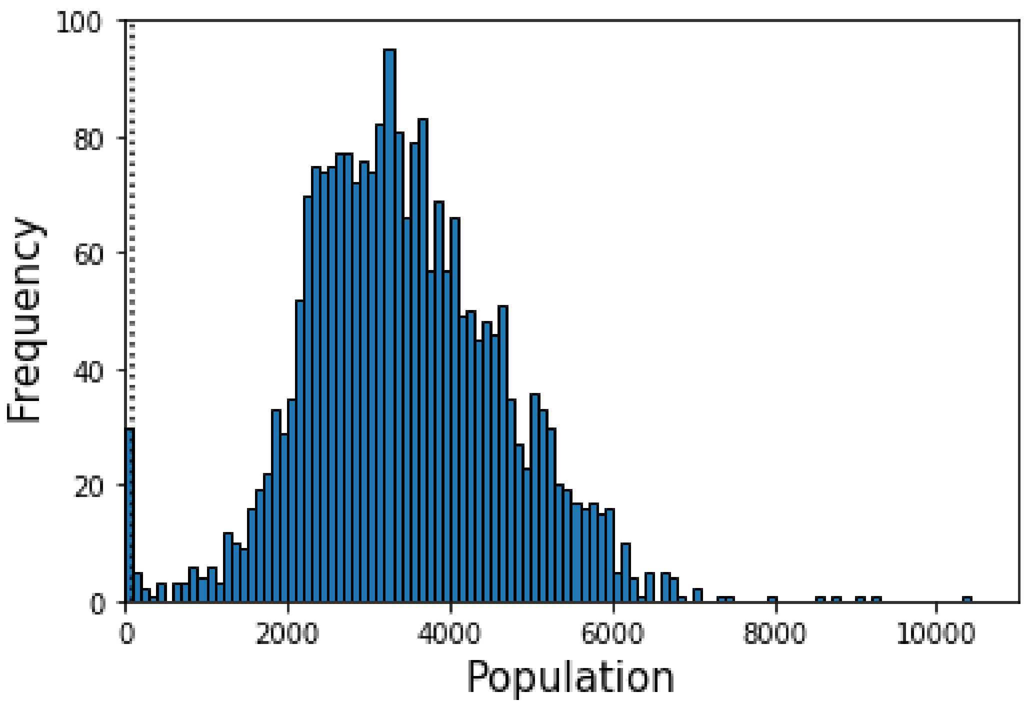
In this specification, New_c indexes whether a tract c contains or will contain a new Metro station within walking distance which is multiplied by $post_t$, an indicator for the post-treatment period after the station is built. $Post_0$ is defined as the time period from 2005-2009 and $post_1$ is defined as the time period from 2012-2016 in order to evaluate a total of three expansions in 2009 and 2012 - the Gold line light rail expansion from Union Station to East Los Angeles, an expansion to the Orange line BRT, and the first leg of the Expo light rail connecting downtown to Culver City. $Existing_c$ indexes whether tract c contains an existing Metro station. Including $Existing_c$ accomplishes two goals. First, it accounts for the network effect in tracts with existing stations in the pre-period, since it is assumed there is a treatment effect of expanding the network in these tracts due to the new Metro stations that increase accessibility across the Metro system. Additionally, this ensures that the appropriate control group is restricted to tracts that do not contain any existing Metro stations. Therefore, two effects are being identified in equation (3.4.2); β_1 identifies the effect of the Metro expansion in tracts that receive a new Metro station relative to tracts that do not have any stations, and β_2 measures the effect of the Metro expansion in tracts that already contain existing Metro stations also relative to tracts that do not have any stations. If a tract is within walking distance of a new Metro station and an existing Metro station, I consider it as *Existing* rather than *New*. The intuition is that although there is a new station in this tract, residents were already connected to the system and the benefit to them is the increased areas for access rather than gaining access to the system, as a new station would provide. Finally, I eliminate tracts that are within walking distance of Metro stations built in the 2016 expansion from the control group.

There are potential shortcomings with using a binary treatment variable. Firstly, if a tract boundary touches the buffer zone, it will be considered as a treated tract. Due to the somewhat irregular shapes of some of the tracts, it is unlikely that the entire tract is within the walking distance buffer – and the differences in the geographical sizes of the tract can be a source of measurement error. More importantly, residents may have different propensities to walk longer distances whereas a fixed walking distance buffer may not be appropriate. For example, higher household incomes and vehicle availability negatively affect a person's propensity to walk (Hsiao et al. 1997). To address this issue, I also employ a continuous distance-based treatment variable that measures the effect of proximity to a Metro transit station on employment in equation (3.4.3).

$$Y_{ct} = \alpha + \beta_1 * KmFromStation_{ct} + \beta_2 * X_{ct} + \Upsilon_c + post_t + \varepsilon_{ct} \quad (3.4.3)$$

$KmFromStation_{ct}$ indexes the distance in kilometers from the centroid of the tract c to the nearest Metro station in time t. As new stations are built, tracts that are near the new station will decrease in their distance to the station. If another station is closer to the tract, then $KmFromStation_{ct}$ will remain unchanged. By using a continuous rather than a discrete measure for Metro stations, I address the issue of residents with different propensities to walk. I express $KmFromStation_{ct}$ using the inverse arcsine to allow for a greater emphasis on changes at smaller distances from a new Metro station. It is important to not use levels, because it is not realistic to expect a constant effect from decreasing distances; a one-kilometer decrease in distance due to a new station has a larger effect in tracts close to Metro stations whereas a one-kilometer decrease from tracts far from the station is not expected to have the same effect. I also express Y_{ct} using the inverse arcsine.[64] [65]

Figure 3.3: Total Tract Population (16+) from 2012-2016



Source: American Community Survey (2012-2016). Bin size is set at 100.

[64] Inverse arcsine is used to approximate a percentage change in place of natural logs because of the presence of tracts with zeros in the outcome variables.

[65] I also test other functional forms, such as levels or 1/KmFromStation. However, since the levels regressions are prone to being affected by outliers and do not account for decreasing distances and the reciprocal form does not have a clean interpretation, the regressions using inverse arcsine are preferred. Results for other functional forms are available upon request.

3.4.2 Instrumental Variables

The construction of a new Metro line may not be exogenous, and may be dependent on a variety of factors, such as existing demand for public transportation, economic indicators, political factors, and geographical practicality. While tract fixed effects assist in controlling for unobserved time-invariant factors, some factors determining Metro placement may vary over time. For example, if there are large changes in existing demand for transportation in a tract, Metro planners may purposely target these areas in their expansion plans. Therefore, using an instrumental variables specification can reduce the bias attributed to endogenous determinants of route placement.

Elkin (2014) provides a history of the decision-making process in rail lines and motivates the creation of an appropriate instrument.[66] As shown in Figure 3.1's timeline, oftentimes routes have pre-determined origin and destination areas that are determined well in advance. The idea of starting with a connection from downtown L.A to Long Beach largely originated from a political compromise on November 21, 1975 between Mayor Bradley and local supervisor Baxter Ward, well before the first construction of the Blue line in 1990. [67] [68] Kenny Hahn, a county supervisor and chair of the Los Angeles County Transportation Commission (LACTC), [69] successfully pushed a proposition for a sales tax hike to fund transit rail plans. Largely due to his individual efforts, the downtown to Long Beach route was prioritized and chosen as the first route in March 1982, which will pass through his predominantly Black, transit-dependent district (Elkin 2014). However, even though the destination was known, this route had resistance to what path it should take. Some wanted a heavy rail route down Vermont Avenue, a busier transit corridor that could service more residents. However, acquiring the funding and political backing would have been more difficult than building on an existing right-of-way owned by Pacific Electric.[70] Even after finally deciding on following the Pacific Electric right-of-way over the Vermont Avenue route, the LACTC still

[66] Elkin (2014) in the book *Railtown: The Fight for the Los Angeles Metro Railway and the Future of the City* provides a succinct historical account of the decision-making process in transit. His primary focus is how the political decisions of rail advocates and detractors and the public support of neighborhoods was the critical factor in determining which routes were prioritized and where they ended up going.

[67] Bradley wished to create a downtown and Hollywood-based subway via Wilshire Blvd, Ward wanted a regional above-ground rail network using existing right of ways (suburban-oriented) (Elkin 2014). These will later develop into the Red/Purple and Blue lines.

[68] After this decision, the federal government organization, the Urban Mass Transportation Administration (UMTA) did not provide funds to the original organization RTD until June 1980 and they insisted on using the funds for a subway system along the Wilshire corridor, closer to Bradley's vision, which will form the basis for the Red Line. A newly-formed organization, LACTC, will secure funding for the Blue Line.

[69] The LACTC and the Southern California Rapid Transit District (RTD) combine into the MTA in 1993

[70] Pacific Electric owned most of the rights-of-way of their former trolley system, the Red Cars.

struggled with the determination of the paths within downtown Los Angeles and downtown Long Beach which were not part of the acquired right-of-way. The plans for the actual route were not finalized until March 1985, nearly a decade after the route was first conceived.

Similarly, for the future Red Line, the Southern California Rapid Transit District (RTD) initially decided in September 1979 on a plan to connect to Hollywood via a route that first goes west to Wilshire and then goes back North and East to Hollywood. However, after a methane gas explosion in 1985, Henry Waxman, the U.S. Representative for the 33rd congressional district, opposed the subway construction through the Wilshire corridor. His concerns were about the risk of another methane gas explosion from the construction; however, he might have also been influenced by the mobilization of his constituency (the wealthy residents of Hancock Park, Beverly Hills, and Fairfax) against the route due to concerns about noise, commuters, and traffic. Therefore, the route had to be changed, and planners decided on an alternative route up through the Vermont corridor to Hollywood and only created a small route down Wilshire, as far as Waxman allowed the route to go.[71]

From as early as 1981-1984, the transportation authorities identified areas where they wished for the routes to be built, including connections from downtown to the San Fernando Valley, Santa Monica, East Los Angeles, and Pasadena (Elkin 2014).[72] Figure 3.4 shows a few of these early proposals and clearly identifies the end destinations of most of the current and future Metro routes. Origins and destination areas are known relatively far in advance, which I treat as exogenous, but the path chosen to connect these areas can be endogenous to factors such as community support and neighborhood demand for transit. To address this endogeneity, I use an instrument to predict where Metro stations are built. This instrument will measure the distance of the centroid of each tract to the nearest point on a hypothetical Metro route connecting the origin and destinations of a specific expansion. Therefore, I am addressing endogeneity from route placement rather than the exact station location along a given route.

---

[71] The Vermont corridor was chosen in part due to limited residents in the area and its composition as more of a business corridor, with a large amount of hospitals.

[72] San Fernando Valley line will later become the Orange Line, the connection to Santa Monica will be the Expo line, and the Pasadena & East LA routes will become the Gold Line

Figure 3.4: Early Planned Routes

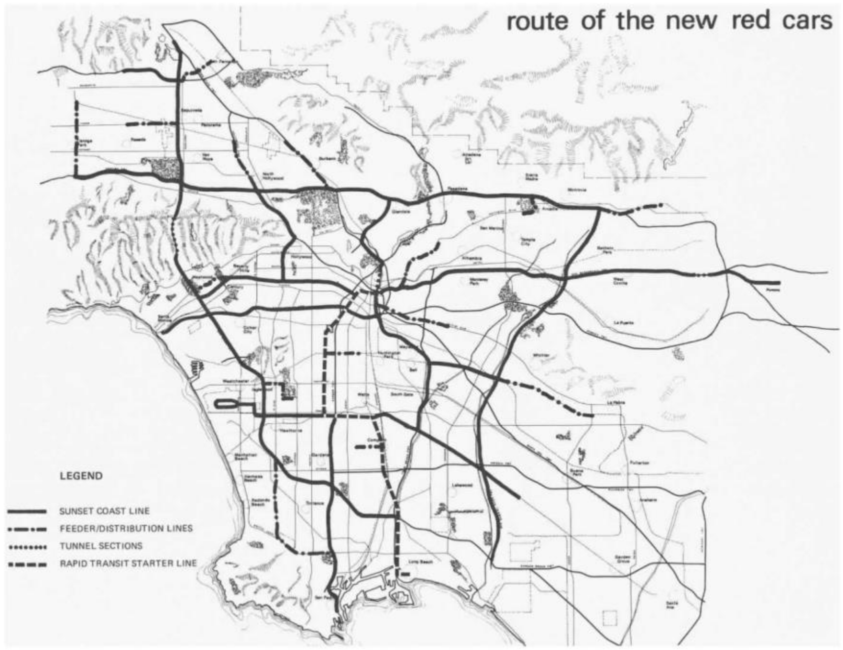


FIGURE 5. Los Angeles county supervisor Baxter Ward's proposed Sunset Coast Rapid Transit Line in 1976. (Courtesy of the Los Angeles County Metropolitan Transportation Authority Research Library and Archive. Reprinted with permission.)

Source: Ballot Proposition A. November 4, 1980

Figure 1·2 Regional Rail Rapid Transit System

Source: Elkin (2014)

I construct a hypothetical Metro route using a least-cost hypothetical route that connects the origin and planned destination by minimizing the changes in the slope of elevation. I constructed this route using a 10 ft digital elevation model (DEM) of Los Angeles County and converted the DEM to a slope raster, which calculates the rate of change in elevation from each cell in the map. The least-cost path is determined by calculating a cost path that minimizes the cost distance, which uses the previous slope raster as the input. Figure 3.5 shows the hypothetical routes that connects the origin and destination points for the Gold line expansion in 2009, phase 1 of the Expo line in 2012, and the expansion of the Orange line in 2012. From this figure, it should be clear that as the distance from any tract c and the least-cost hypothetical route increases, the probability of a Metro station being constructed in tract c decreases. Conversely, when using the continuous treatment variable, the kilometers from the nearest Metro station, as the distance to the hypothetical route increases, so should the distance to the nearest actual Metro station. There are a few usages of constructing least-cost hypotheticals to use as instruments. Lipscom et al. (2013) and Kaseem (2018) construct a hypothetical electrical grid in the countries they study based on geographical-based cost considerations. Neumark (2008) uses distance to Benton County to predict the opening of new Walmart stores. The benefit of this type of instrument is that along with a strong first stage, the theoretical argument for the exclusion principle is strong. It is unlikely that the distance of the centroid of each tract to a hypothetical Metro route affects labor market outcomes except through its prediction of actual stations.

More formally, the first-stage IV specification relating to equation (3.4.2) is as follows:

$$New_c * post_t = \alpha + \delta * Hypo\_Route_c * post_t + \lambda_1 * Existing_c * post_t + \lambda_2 X_{ct} + \Upsilon_c + post_t + \varepsilon_{ct} \qquad (3.4.4)$$

$Hypo\_Route_c$ is a vector of distance dummies that indicate whether the distance from the centroid of tract c is less than 1km, 2km, or 3km from the tract centroid to the hypothetical route.[73] Note that $Hypo\_Route_c$ does not vary with time, so I will interact it with $post_t$ or the instrument will be absorbed by the fixed effects, and the distance is not relevant for new lines until they are constructed in the post period.

Additionally, an equivalent distance-based IV first-stage modification of equation (3.4.4) would be as follows:

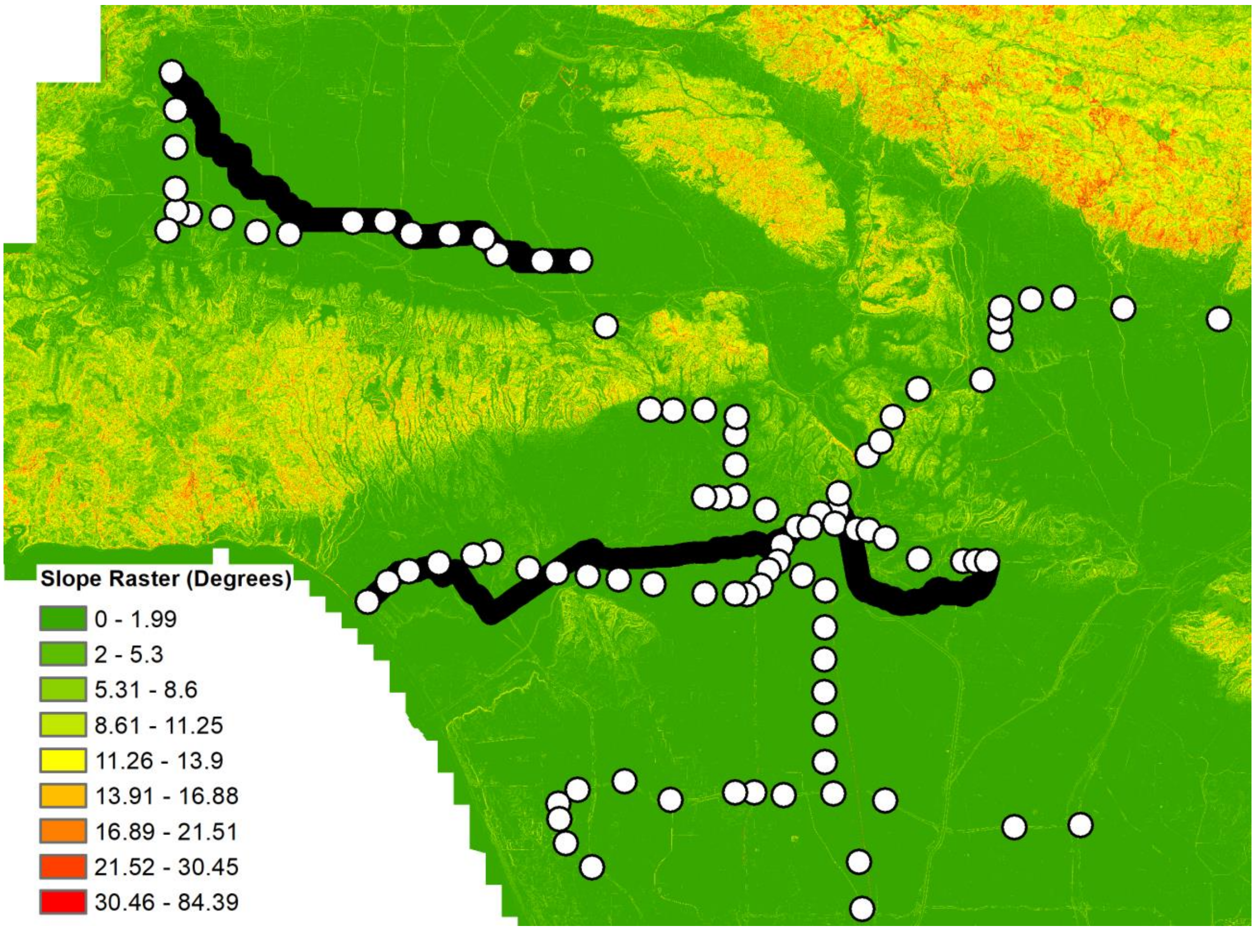
[73] When the treatment variable is binary, as is the case here, the instrument cannot be continuous or there will be a forbidden regression problem (Angrist and Pischke 2009)

$$KmFromStation_{ct} = \alpha + \delta_1 * Hypo\_Route_c * post_t + \delta_2 X_{ct} + \Upsilon_c + post_t + \varepsilon_{ct}$$

(3.4.4)

Here, $Hypo\_Route_c$ is simply the continuous distance from the centroid of the tract to the hypothetical route. I expect δ1 to be positive here – as the distance to the hypothetical line increases, so does the distance to the actual Metro station.

Figure 3.5: Hypothetical Least-Cost Route for Origin-Destination



Source: Los Angeles County GIS Data Portal. The digital elevation model is converted to a slope raster (shown in the background) and the least-cost route is calculated using spatial analysis tools in ArcMap that account for elevational slope changes.

One potential criticism of the least-cost route is that the hypothetical route mapped out is not feasible without acquiring the rights-of-way. Metro must acquire the right-of-way to build a transit route on the land, and it is much easier to acquire these rights if they were previous streetcar or railroad lines in the past. For example, in 1990, Metro purchased 175 miles of rights-of-way for $450 million from Southern Pacific, including the 21-mile Burbank Branch line in San Fernando

Valley (the future Orange line) and the 14-mile Exposition Boulevard route from downtown to Santa Monica (the future Expo line). Also, in 1990, the LACTC purchased the right-of-way for the Santa Fe railroad tracks to Pasadena (future Gold line). Thus, Metro builds on routes that have clearly defined and owned rights-of-way that are primarily based on the previous streetcar systems. The streetcar system appears to form the basis for most routes, with the exception of the Gold Line which uses the Santa Fe Railroad tracks and the extensions of some of the other routes, such as the Green line which uses a freeway median[74] for most of the route and the last two stations in the southern portion of the Gold line that connects through East Los Angeles.[75] Therefore, I proxy for the universe of all rights-of-way that Metro could have conceivable acquired by mapping out the historical streetcar routes of Pacific Electric and the Los Angeles Railway as well as existing railroads in the United States (see Figure 3.6 Panel A).[76]

One obvious potential instrument using the historical routes is to construct the hypothetical route by isolating the possible paths that connect the origin and destination areas with the historical rights-of-way. For example, to connect to Santa Monica in the West, it looks like there are multiple viable paths. Using historical routes as instruments is supported in the transportation literature to study transportation-related outcomes such as modern traffic patterns (Redding and Turner 2015; Duranton and Turner 2011). However, there are a few reasons why this is not the appropriate method to use in this context. First, and most importantly, IV is only appropriate if it meets the exclusion restriction. However, in this context, the distance of a tract to these early historical rights-of-way routes may affect contemporaneous labor market conditions, violating the exclusion restriction. It is conceivable that areas close to early historical railroads or the streetcar system led to economic development that persist over time and affect current labor market conditions. Brooks and Lutz (2019) demonstrate the persistence of the LA streetcar system on population densities,

---

[74] Interestingly, most of the Green line, before it veers south on the west end, is probably the best example of an exogenously determined route for the portion on Interstate 105. It was the result of a court settlement regarding acquisition of land for Century Freeway (Interstate 105) that required a stipulation requiring a median be built for transit on the highway. However, beyond the freeway, the Green line did not reach its goals. It failed to connect to the Norwalk Amtrak station due to opposition from Norwalk leaders, ran into limitations in connecting to LAX directly due to bureaucracy at the airport and ended up with a station 2 miles from LAX, and decided to go south to serve the Redondo Beach aerospace industry (which ultimately was severely crippled by the end of the Cold War when it was finally completed) (Elkin 2014)

[75] While the Orange line looks like it differs from the actual route placement, I attribute this to mapping error. The Orange line is supposed to be based on the Pacific Electric (Red Cars) route.

[76] It is assumed that all existing railroads are also based on historical rights-of-way. For example, the Gold line is based off the Santa Fe Railway built in 1859. These are primarily freight routes, and it can be assumed that Metro could have conceivable purchased the rights-of-way to convert to commuter lines.

attributable to land use regulation, which lends credence to this idea of economic persistence. In addition to the exclusion restriction issue, it is not clear which historical routes should be chosen in constructing the hypothetical route that connects the origin and destination points – whether it should be the shortest connection, the one that services the most residents, or any possible route not longer than a certain threshold. Choosing too many potential routes leads to a weak first-stage. Therefore, I prefer to use the least-cost route as the instrument since it yields a stronger first stage and satisfies the exclusion principle.

However, these historical routes can serve as an important robustness check. The idea of persistence of higher economic growth in neighborhoods that are close to these historical routes indicates that the IV estimates may be upward biased when using all tracts without Metro stations in Los Angeles County as the control group. Therefore, the control group should be restricted to tracts that are close to the historical streetcar system or railroads, which I defined as a 0.8-kilometer buffer region (see Figure 3.6 Panel B).[77] I test this theory by determining whether there is a spurious effect by comparing the outcomes of tracts without Metro stations but part of the historical right-of-way against tracts without Metro stations and not part of a historical right-of-way with the following equation:

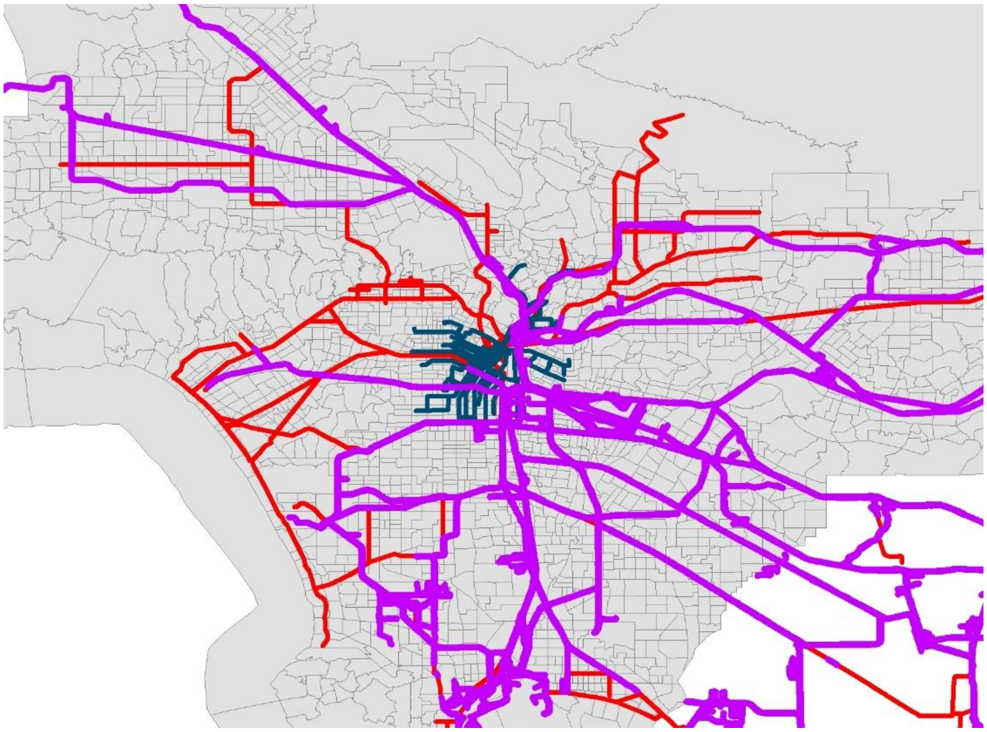
$$Y_{ct} = \alpha + \beta_1 * Historical_c * post_t + \beta_2 * X_{ct} + \Upsilon_c + post_t + \varepsilon_{ct} \qquad (3.4.5)$$

In this specification, $Historical_c$ indicates whether a tract is within a 0.8-kilometer walking distance buffer around the historical streetcar route or railroad. All tracts that will contain Metro stations within walking distance are omitted. If the theory is true, then β_1 should generate a positive, spurious effect.

[77] I only conduct this robustness check with the binary specification of equation (2.3.2) and equation (2.3.4), because I also make the 0.8-kilometer buffer region assumption in those specifications. It is not appropriate in the continuous specification of the treatment variable – although one might consider assigning different weights to tracts depending on their proximity to these historical routes.

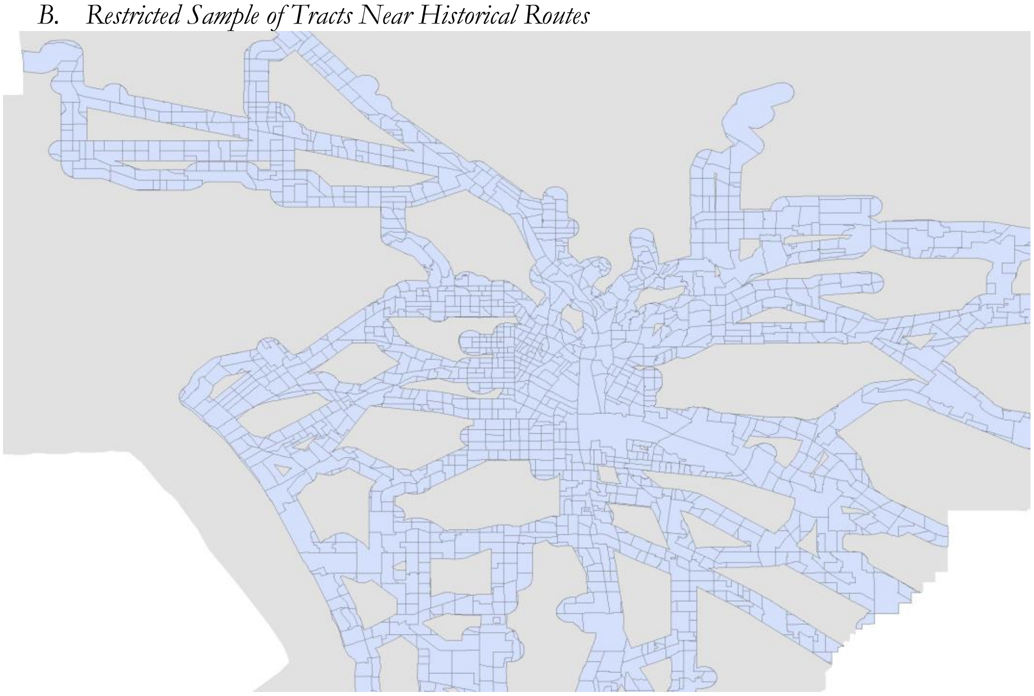
Figure 3.6: Rights-of-Way Based on Historical Routes

A. Red Cars, Yellow Cars, and Currently Operating Railroads



The red routes indicate a Pacific Electric streetcar route (Red cars). The blue routes indicate a Los Angeles Railway route (Yellow cars). The purple indicates a currently operating railroad.

B. Restricted Sample of Tracts Near Historical Routes



This shows a map of tracts that are counted in the restricted subsample which is limited to a 0.8-kilometer walking distance around the routes above.

3.5 Results

Figure 3.7 compares a descriptive map of labor force participation in an area of tracts where most of the evaluated expansions fall under. In this figure, the treated areas contained within the circle have blue or black outline, which indicates they are part of the 2009 and 2012 expansions. A few things stand out from the figure. First, labor force participation appears to be lower in the treated areas in the pre-period (2005-2009) with most tracts in these areas not having labor force participation above 75%, except for the western-most station. Taking an aggregate of all treated and control tracts, the average LFP of the control tracts rate in the pre-period is 65% while the average LFP in the treated tracts is 62%.[78] Thus, it is conceivable that planners target neighborhoods with lower LFP. Additionally, it does appear that labor force participation is increasing in the post-period in many of the tracts – this is more evident from the stations within the 2009 expansions than the 2012 expansions. In the post period, LFP of control tracts decreases by one percentage point while the average of treated tracts increases by one percentage point. However, these are just averages - the regression estimates will provide a better estimate of the effect of Metro stations on labor force participation and other measures.

Table 3.1 Panel I shows the regression results matching equation (3.4.2). Starting with public transit usage, the estimate shows an increase of rail or subway usage by 0.35 percentage points in tracts where the new stations are placed and a 0.40 percentage point increase in tracts with existing Metro stations. The increased rail and subway usage are offset by the decrease of 0.65 percentage points in bus usage in tracts with new stations. The corresponding IV estimate in Panel II indicates that the initial estimate is understated, and the IV estimates suggest a 0.52 percentage point increase in subway and rail usage and decrease in 1.16 percentage points in bus usage, although the latter is not statistically significant. There are a few points when interpreting these results. First, the Orange bus rapid transit system expanded four stations in 2012, so the effect cannot be completely attributed to workers substituting out buses for the light rail systems. Nevertheless, given that the Gold and Expo lines expanded a total of 17 light rail stations compared to 4 stations, it is still safe to infer that there is a large degree of substitution of workers who previously used buses switching to the new light rail system. Secondly, note that bus usage decreases at a higher rate than the increase in rail usage. While not shown in the table, the overall effect on public transit share is close to the sum

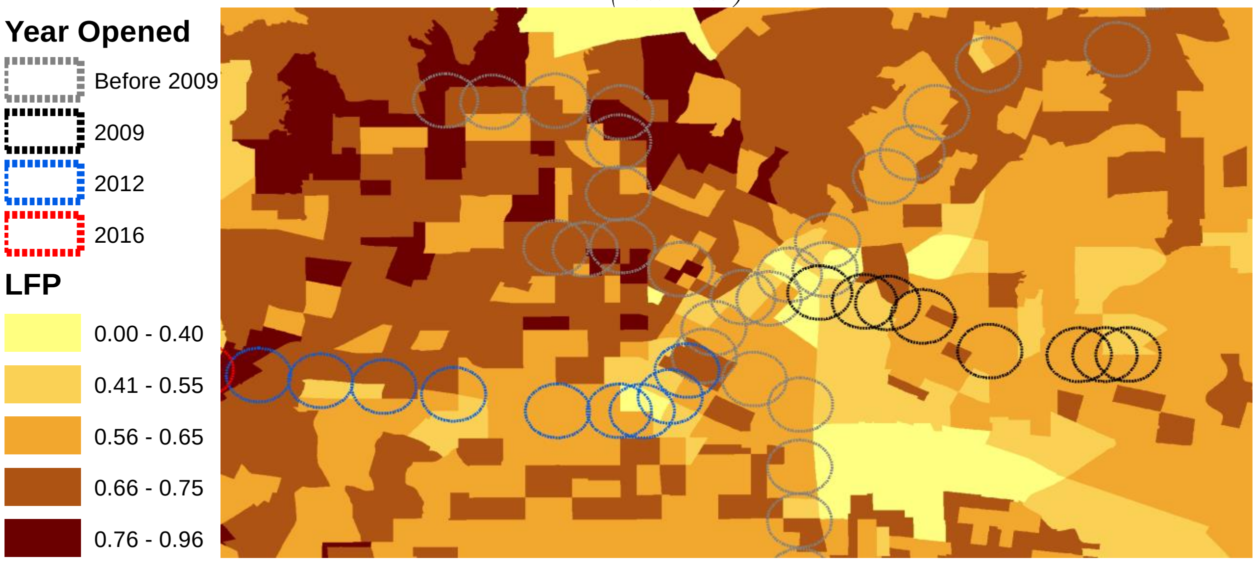
---

[78] The employment to population rate is 60% for non-treated tracts and 56% for treated tracts. The employment to population figures are similar to LFP, so I omitted them.

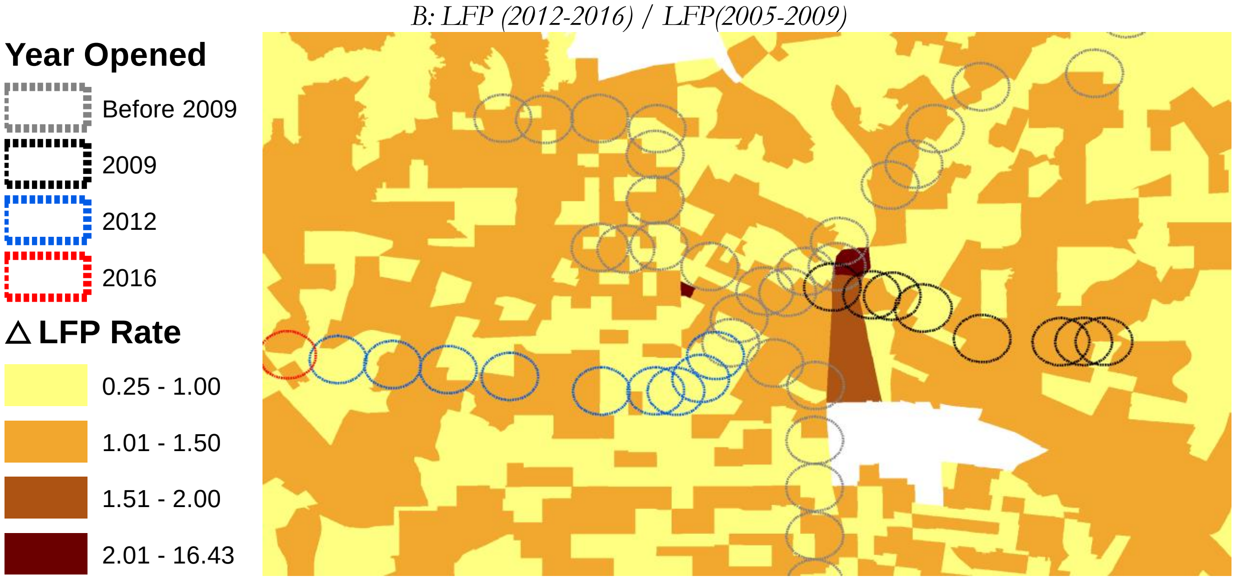
of both the subway/rail usage and the bus usage estimates, which is negative but not statistically significant under both the DiD/IV specifications for tracts with new stations, although statistically significant for tracts with existing stations. Thus, the overall public transit share is only increasing in tracts with existing stations while in the tracts with new stations, existing workers are substituting bus for rail.

Figure 3.7: Labor Force Participation in 2005-2009 Compared to 2012-2016

A: LFP (2005-2009)



B: LFP (2012-2016) / LFP(2005-2009)



The circles represent a 0.8-kilometer buffer region around a Metro station. The black circle is part of the 2009 expansions, the blue circles are part of the 2012 expansions, and the red circles are part of the 2016 expansions (which is omitted in this analysis). The silver circles are from existing stations built in 1990-2005. Both LFP (2005-2009) and LFP (2012-2016) use the same scale. This figure is zoomed in to the areas near downtown and where most of the expansions occurred. Tract boundaries are not shown.

Table 3.1: The Effect of Metro Stations on Labor Force Participation, Commuting, and Earnings

Variables	(1) Subway or Rail	(2) Bus	(3) LFP (All)	(4) LFP (Black)	(5) LFP (Hisp)	(6) LFP (< HS)	(7) Emp (All)	(8) Emp (Black)	(9) Emp (Hisp)	(10) Emp (< HS)	(11) Yearly Earnings
I. DiD											
New*post	0.35***	-0.67**	1.80***	4.22**	1.85**	0.06	1.85***	4.64***	2.04**	0.50	-1044.48*
	[0.08]	[0.29]	[0.58]	[1.70]	[0.80]	[1.28]	[0.57]	[1.61]	[0.89]	[1.37]	[591.04]
Existing*post	0.40***	-0.13	1.99***	3.13***	1.81***	1.82**	2.04***	3.42***	2.09***	2.35***	229.59
	[0.06]	[0.18]	[0.35]	[1.12]	[0.47]	[0.71]	[0.36]	[1.15]	[0.54]	[0.73]	[348.19]
constant	0.33	5.14***	64.10***	55.10***	69.33***	72.41***	58.55***	42.08***	66.55***	71.78***	26561.66***
	[0.27]	[1.03]	[2.40]	[6.88]	[2.93]	[3.55]	[2.35]	[6.95]	[3.22]	[3.72]	[2422.45]
II. IV											
New*post	0.52**	-1.16	7.53***	6.79**	5.17***	1.45	6.51***	9.81***	4.04**	-0.39	-3863.48**
	[0.22]	[0.83]	[1.74]	[3.01]	[1.74]	[2.73]	[1.67]	[3.15]	[1.97]	[2.91]	[1746.26]
Existing*post	0.41***	-0.15	2.22***	3.31***	2.00***	1.91***	2.23***	3.78***	2.20***	2.29***	116.40
	[0.06]	[0.19]	[0.35]	[1.14]	[0.47]	[0.70]	[0.36]	[1.15]	[0.55]	[0.73]	[355.92]
First-Stage F-Stat	31.38	31.38	31.38	18.01	29.78	29.04	31.38	18.01	25.46	29.04	31.37
Endogeneity Test (p-value)	0.337	0.435	0.000	0.230	0.021	0.874	0.000	0.062	0.226	0.752	0.088
Overidentification Test (p-value)	0.070	0.083	0.307	0.778	0.632	0.160	0.136	0.524	0.528	0.423	0.912
Observations	4508	4508	4508	3936	4504	4436	4508	3936	4152	4436	4502

Source: American Community Survey (ACS) 5-year summary files, 2005-2009 ACS and 2012-2016 summary ACS files. This table reports DiD/IV estimates of equation (3.4.2) using the 0.8-kilometer walking distance buffer around Metro stations. Tracts that had expansions in 2009 and 2012 are considered *New* while tracts that have existing stations from 1990-2005 are considered *Existing*. Tracts with expansions in 2016 are omitted from the regression. In column (1), (2), (6), and (10), the specification includes controls for shares female, Black, Asian, two or more races, Hispanic, high school degree, some college, bachelor's degree or higher, out of county migration in the past year, out of state migration in the past year, no vehicles, and one vehicle. Column (3), (4), (7), and (8) contain the same specification as (1) but exclude shares Black, Asian, two or more races, and Hispanic. Columns (5) and (9) contains the same specifications as (1) except without shares with high school degree, some college, and bachelor's degree or higher. All columns include tract fixed effects and a post period dummy variable. Regression is weighted by their respective population (i.e. all population 16 and over for LFP (All), teen population for LFP (teen), etc. The first stage F-Stat differs due to the sample size data arising from censored data in omitted tracts for subgroups. The units for columns (1)-(10) represent percentage points from 0-100, and the units for column (11) is in US dollars. Standard errors are clustered by tract. *, **, and *** denote 0.1, 0.05, and 0.01 significance levels.

If I expect labor force participation and employment only increases because more workers are utilizing the increased public infrastructure to commute and search in distant locations, then from the previous estimates, there should not be a significant effect on labor force participation or employment. However, the effects on the labor force participation rate and the employment to population ratio are strikingly clear. Initial DiD estimates indicate a statistically significant increase in the labor force participation rate by 1.80 percentage points and employment to population rate by 1.85 percentage points for tracts with new stations within walking distance. There appears to be a much larger effect on Blacks, with a 4.22 percentage point increase in the labor force participation rate and a 4.64 percentage point increase in the employment rate. The estimates for Hispanics are similar to the estimates for the general population, and not significant for those without a high school degree. The secondary effects for tracts with existing stations are all positive and significant for all groups, ranging from 1.81 to 3.13 percentage points for labor force participation and 2.04 and 3.43 percentage point increases for employment.

IV estimates suggest that the initial estimates are understated, with an increase in labor force participation by 7.53 percentage points and employment by 6.51 percentage points. Among the subgroups, IV estimates suggest that the effect on labor force participation of Blacks (an increase in 6.79 percentage points) and Hispanics (an increase in 5.17 percentage points) are smaller than on the effects on the overall labor force participation rate (an increase in 7.53 percentage points). However, there appears to be a larger effect on employment for Blacks (an increase in 9.81 percentage points) than the general employment rate (an increase in 6.51 percentage points). While the magnitude of the IV estimates may be slightly inconsistent with the magnitude of the initial estimates among the different subgroups, such as the estimate for Black LFP being lower than the overall LFP, this might be explained by the differing strength of the first stage. The first stage regression specification is the same but the sample sizes differ due to missing data for a number of tracts where the population of the subgroup is small, leading to a weaker first stage F-Statistic, 18.01, when compared to the F-statistic using the full sample, 31.38. For context, the Stock-Yogo critical values for the 5% maximal IV relative bias is 13.91 and for 10% maximal IV size is 22.30 (although for 15% maximal IV size is 12.83). Additionally, the test of endogeneity is significant for labor force participation and employment for the overall population and Hispanics, as well as Black employment, but not significant for Black LFP and unskilled LFP and employment.

On earnings, the initial DiD estimate indicates a decrease in $1,044 in yearly earnings for the average resident in the treated tract, statistically significant at the 10% level. The corresponding IV estimate suggests the magnitude of this decrease is understated, with estimates indicating a decrease of $3,460 in yearly earnings. This is not necessarily contradictory to employment estimates – as it depends on the type of jobs that residents are getting employed into. Residents may be getting employed into lower-pay and/or part-time jobs, which is consistent with the idea that lower income residents tend to rely on public transit or vulnerable to spatial mismatch when compared to more than higher income residents. Confidential data could paint a better picture with the distributional effects, particularly for the different subgroups being studied.

Recall that the persistence of economic conditions in neighborhoods close to historical routes can generate bias, as supported by Brooks and Lutz (2019). I conducted a robustness test and restricted the control group to only tracts that are within walking distance of historical routes whose rights-of-way could have been purchased by Metro to convert into Metro routes. I first test that this is an appropriate analysis by comparing the outcomes variables between the new control group, tracts near historical routes, versus the initial control tracts. In Table 3.2 Panel I, I find a statistically significant and positive spurious effect in labor force participation and employment (except for Hispanics). This provides evidence that restricting the control group to tracts near historical routes is a more credible control group and these estimates will be preferred over Table 3.1.

After using the restricted subsample of tracts, I find weaker effects for labor force participation and employment with the new control group, as expected, but most of the estimates are still statistically significant. For example, initial DiD estimates indicate a 1.42 percentage point increase in the labor force participation rate and IV estimates indicate a 5.49 percentage point increase (1.52 and 4.93 percentage point increases for employment). For Blacks, both DiD/IV estimates suggest smaller effects than Table 3.1. Initial estimates indicate a 3.5 percentage point increase for LFP and a 3.9 percentage point increase for employment, while IV estimates indicate a 4.51 percentage point increase for LFP and a 7.32 percentage point increase for employment, although the IV estimate for LFP is no longer significant. The effects on public transit, LFP for Hispanics, employment for Hispanics, and earnings do not differ much from the estimates in Table 3.1, nor were they expected to, since there was not a spurious correlation between the tracts near historical routes compared to the tracts without historical routes. The first-stage F statistics are slightly stronger here, ranging from 19.42 to 32.54 depending on the sample size.

Table 3.2: Effect of Metro Stations on Labor Force Participation, Commuting, and Earnings (Restricted to Historical Routes)

Variables	(1) Subway or Rail	(2) Bus	(3) LFP (All)	(4) LFP (Black)	(5) LFP (Hisp)	(6) LFP (< HS)	(7) Emp (All)	(8) Emp (Black)	(9) Emp (Hisp)	(10) Emp (< HS)	(11) Earnings
I. Historical Route Test											
Historical*post	-0.01	-0.05	1.24***	2.77**	0.52	1.63**	0.97***	2.60**	0.20	1.81**	361.12
	[0.03]	[0.09]	[0.29]	[1.08]	[0.48]	[0.76]	[0.28]	[1.07]	[0.51]	[0.77]	[390.30]
constant	0.47	4.73***	64.41***	47.57***	72.81***	75.49***	57.78***	29.80***	67.32***	72.32***	27751.68***
	[0.29]	[1.04]	[2.65]	[8.22]	[3.36]	[4.37]	[2.53]	[8.03]	[3.54]	[4.52]	[2965.44]
Observations	3600	3600	3600	3094	3596	3538	3600	3094	3354	3538	3596
II. DiD - Main Specification											
New*post	0.36***	-0.66**	1.42**	3.48**	1.74**	-0.32	1.52***	3.86**	2.00**	0.07	-1164.38*
	[0.08]	[0.29]	[0.59]	[1.71]	[0.81]	[1.31]	[0.58]	[1.64]	[0.90]	[1.39]	[597.68]
Existing*post	0.40***	-0.11	1.64***	2.23*	1.71***	1.47**	1.76***	2.57**	2.03***	1.95***	143.64
	[0.06]	[0.19]	[0.36]	[1.19]	[0.48]	[0.73]	[0.37]	[1.20]	[0.55]	[0.75]	[355.57]
constant	0.42	5.51***	66.27***	51.59***	71.23***	72.17***	61.78***	39.22***	70.23***	73.64***	24205.32***
	[0.30]	[1.22]	[2.67]	[7.67]	[3.18]	[3.81]	[2.62]	[7.92]	[3.49]	[3.98]	[2502.15]
III. IV - Main Specification											
New*post	0.48**	-0.94	5.49***	4.51	4.88***	-0.00	4.93***	7.32**	4.31**	-0.90	-3634.59**
	[0.21]	[0.83]	[1.63]	[3.07]	[1.72]	[2.75]	[1.60]	[3.11]	[1.95]	[2.92]	[1634.36]
Existing*post	0.41***	-0.13	1.87***	2.33*	1.93***	1.50**	1.95***	2.92**	2.18***	1.87**	5.78
	[0.06]	[0.19]	[0.37]	[1.22]	[0.49]	[0.73]	[0.38]	[1.22]	[0.56]	[0.75]	[366.85]
First-Stage F-Stat	32.54	32.54	32.54	19.42	30.66	29.53	32.54	19.42	26.36	29.53	32.53
Endogeneity Test (p-value)	0.448	0.626	0.003	0.544	0.027	0.978	0.009	0.200	0.131	0.748	0.112
Overidentification Test (p-value)	0.129	0.074	0.639	0.490	0.580	0.185	0.263	0.177	0.380	0.349	0.843
Observations	3458	3458	3458	3016	3456	3416	3458	3016	3160	3416	3452

Source: *I.* follows the specification in equation (3.4.5). *Historical* indexes whether a tract is within walking distance (0.8-kilometers) of a historical route or current railroad route. It uses the same control variables and weights as Table 3.1. *II.* and *III.* follow the same specifications as the regressions in Table 3.1, except restricting the control group to tracts where *Historical* = 1. The observations in the last row apply to both *II.* and *III.* The units for columns (1)-(9) represent percentage points from 0-100, and the units for column (10) is in US dollars. Standard errors are clustered by tract. *, **, and *** denote 0.1, 0.05, and 0.01 significance levels.

Table 3.3: Effect of Proximity to Metro Stations on Labor Force Participation, Commuting, and Earnings

Variables	(1) Subway or Rail	(2) Bus	(3) LFP (All)	(4) LFP (Black)	(5) LFP (Hisp)	(6) LFP (< HS)	(7) Emp (All)	(8) Emp (Black)	(9) Emp (Hisp)	(10) Emp (< HS)	(11) Earnings
1. DiD											
KmFromStation	-0.136***	0.075**	-0.009*	0.005	-0.014**	-0.004	-0.009	-0.003	-0.016*	0.003	-0.004
	[0.029]	[0.030]	[0.005]	[0.019]	[0.007]	[0.011]	[0.006]	[0.020]	[0.008]	[0.014]	[0.009]
constant	0.564***	1.566***	4.906***	4.689***	4.969***	4.918***	4.825***	4.485***	4.938***	4.922***	10.701***
	[0.207]	[0.262]	[0.053]	[0.153]	[0.064]	[0.070]	[0.057]	[0.178]	[0.079]	[0.080]	[0.067]
II. IV											
KmFromStation	-0.401***	-0.172	-0.199***	-0.273***	-0.190***	-0.148***	-0.186***	-0.246***	-0.171***	-0.156***	0.021
	[0.091]	[0.121]	[0.025]	[0.064]	[0.032]	[0.044]	[0.026]	[0.067]	[0.040]	[0.051]	[0.031]
III. First Stage Equation (KmFromStation on Hypo Route Distance)											
Hypo Route Distance	0.010***	0.010***	0.009***	0.013***	0.010***	0.010***	0.009***	0.013***	0.009***	0.010***	0.009***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
First-Stage F Stat	203.76	203.76	203.76	87.25	136.76	123.09	203.76	87.25	129.43	123.09	203.38
Endogeneity Test (p-value)	0.003	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.395
Observations	4604	4604	4604	4028	4600	4532	4604	4028	4248	4532	4598

Source: American Community Survey (ACS) 5-year summary files, 2005-2009 ACS and 2012-2016 summary ACS files. This table reports DiD/IV estimates of equation (3.4.3) using a continuous treatment variable, *KmFromStation*, which measures how far the centroid of the tract is from the nearest Metro station, expressed in its inverse arcsine form. The outcome variables in columns (1)-(10) are also expressed in inverse arcsines. See Table 3.1 for description of control variables. All columns include tract fixed effects and a post period dummy variable. Standard errors are clustered by tract. The pre-period defined here is 2005-2009 and the post period is 2012-2016.The first stage estimates differ due to sample size differences from censored data in omitted tracts for subgroups. *, **, and *** denote 0.1, 0.05, and 0.01 significance levels.

Additionally, I test the effects of a continuous distance to a Metro station on the above outcomes of interest, matching with equation (3.4.3), shown in Table 3.3.[79] In the base specification using $KmFromStation_{ct}$, I expect the sign of the estimate on $KmFromStation_{ct}$ to be negative, because the positive effect of Metro stations should be decreasing as distance from the station increases. Therefore, I will be interpreting the effects in this section as an increase in *proximity* rather than distance, to make the interpretation of the estimates easier. For example, the IV estimate in the Table 3.3 suggests a one percentage increase in *proximity* to a Metro station increased subway or rail usage by 0.401 percent on average. The specification using the continuous distance measure also has the benefit of a stronger first stage – with a F-stat of 203.76 for the full sample.[80]

Table 3.3 shows that subway or rail usage, labor force participation, Hispanic labor force participation, and Hispanic employment increases with increased proximity to stations, while bus usage decrease with increased proximity to stations. IV estimates in the distance-based specification are mostly consistent with the direction of the bias expected. For labor force participation, the IV estimates can be interpreted as for each one percent increase in *proximity* to the Metro station, labor force participation rate on average increases by 0.199 percent. The effects are largest for Black residents, who increase their labor force participation rate by 0.273 percent on average for each one percent increase in proximity. Similarly, the effect of proximity to station on employment is smaller in magnitude than labor force participation but follows the same pattern.

In general, IV estimates using the continuous specification are consistent with the binary specification, with a few exceptions. The estimates for labor force participation and employment are significant for high school dropouts in the continuous case, but not the binary case. Additionally, the IV estimates for bus usage and earnings are not significant using the continuous specification. Overall, the IV may perform better in the continuous distance specification rather than the binary case, because the equation is just identified and the first stage F-statistics are much stronger, ranging from 87.25 to 203.76. Finally, the endogeneity test rejects the null that KmFromStation is exogenous at the 1% level, except in the earnings regression. I do not apply the restriction of historical tracts for a robustness check since I assumed of a 0.8-kilometer walking buffer around a historical streetcar

[79] I do not eliminate tracts where the Metro station is built in 2016 in the continuous specification, because this requires an assumption about walking distances that was used in the discrete case. Nevertheless, I do test the results when eliminating tracts using the walking distance assumption and conclude that the results do not change by much.

[80] Again, the first stage estimates are slightly different between each column due to different weights and control variables being used (i.e for the estimates of the labor force participation of Blacks, I use Black population as weights and omit race controls).

to be consistent with how I constructed the binary measures of New_c and $Existing_c$ around Metro stations.[81]

Overall, I have shown that there is robust evidence that Metro stations have a positive effect on labor force participation and employment for residents in the neighborhoods these stations are placed in. However, the lack of a strong uptick in public transit usage suggests that workers in these neighborhoods who were reliant on public transit were merely substituting between different types of transportation methods rather than being connected to distant jobs. If new workers who were at the extensive margin choose to participate in the labor force due to accessibility to distant jobs, the overall public transit share should increase at the same magnitude as employment. Thus, neighborhood changes may provide a better explanation. If the explanation is that Metro stations increased economic activity in neighborhoods near the stations, then the increased labor force participation and employment is a direct result of increased job density near the station.

I turn to another the publicly available Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) in the absence of confidential ACS data on place of work to determine if proximity to Metro stations affects job density. Table 3.4 shows the results using the binary specification of tracts within walking distance (eq 3.4.2). I do not find any statistically significant results on number of jobs or number of jobs per resident population in the initial DiD estimates although most estimates are positive, except for medium income jobs and medium income jobs per population (Table 3.4 Panel I). There also is a sizable and statistically significant (at the 10% level) increase in high income jobs in tracts that have existing stations. The IV results are also not statistically significant, except for a decrease in low income jobs per capita. However, commuting distance decreases, with initial DiD estimates indicating an average drop of 0.38 kilometers and IV estimates indicating a decrease in 2.71 kilometers in tracts with new stations, evidence that workers are now employed in jobs closer to their homes. Restricting to historical tracts (Table 3.5) yields similar, inconclusive results except for decreases in average commuting distances.

[81] I may consider applying different weights based on the distance of the tract from the nearest historical route, but that has not been implemented.

Table 3.4: Effects of Metro Stations on Jobs (Binary Specification)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Data Source	ACS	LODES	LODES	LODES	LODES	Both	Both	Both	Both	LODES
Variables	Resident Population	Total Jobs	Low Inc. Jobs	Med Inc. Jobs	High Inc. Jobs	Total Jobs / Pop	Low Inc. Jobs / Pop	Med Inc. Jobs / Pop	High Inc. Jobs / Pop	Commute Distance
I. OLS										
New*post	-2.58	84.79	34.09	-12.66	63.36	0.03	0.01	-0.01	0.02	-0.38***
	[40.14]	[107.06]	[52.04]	[36.34]	[49.32]	[0.04]	[0.02]	[0.02]	[0.02]	[0.07]
Existing*post	23.18	119.18	2.43	-21.11	137.86*	-0.05	-0.03	-0.05**	0.03	-0.32***
	[28.83]	[123.07]	[50.37]	[31.14]	[81.14]	[0.10]	[0.02]	[0.02]	[0.08]	[0.07]
constant	3715.78***	1904.35***	415.35***	721.65***	767.35**	1.06***	0.17***	0.30***	0.59	18.04***
	[190.82]	[443.94]	[142.77]	[106.23]	[380.59]	[0.41]	[0.05]	[0.06]	[0.39]	[0.49]
II. IV										
New*post	-132.81	316.63	-154.87	18.18	453.33	-0.06	-0.13*	-0.11	0.18	-2.71***
	[139.56]	[447.04]	[149.90]	[123.75]	[307.81]	[0.40]	[0.08]	[0.10]	[0.31]	[0.40]
Existing*post	17.62	128.48	-5.15	-19.87	153.51*	-0.06	-0.03	-0.06*	0.03	-0.42***
	[30.12]	[133.30]	[49.19]	[34.58]	[90.75]	[0.12]	[0.02]	[0.03]	[0.09]	[0.07]
First Stage F Stat	33.30	31.38	31.38	31.38	31.38	31.38	31.38	31.38	31.38	31.38
Endogeneity Test (p-value)	0.401	0.467	0.892	0.512	0.444	0.726	0.069	0.658	0.304	0.000
Overidentification Test (p-value)	0.415	0.311	0.165	0.942	0.116	0.691	0.515	0.351	0.395	0.418
Observations	4508	4508	4508	4508	4508	4508	4508	4508	4508	4508

Source: Longitudinal Employer-Household Origin-Destination Employer Statistics (LODES) and ACS data. LODES data aggregated to tract level to match ACS specifications. Low-income jobs have earnings of $1250/mo or less, medium-income jobs have earnings of $1251/mo to $3333/mo, and high-income jobs have earnings with greater than $3333/mo. Robust standard errors clustered by tracts. *, **, and *** denote 0.1, 0.05, and 0.01 significance levels.

Table 3.5: Effects of Metro Stations on Jobs (Restricted to Historical Routes)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Data Source	ACS	LODES	LODES	LODES	LODES	Both	Both	Both	Both	LODES
Variables	Resident Population	Total Jobs	Low Inc. Jobs	Med Inc. Jobs	High Inc. Jobs	Total Jobs / Pop	Low Inc. Jobs / Pop	Med Inc. Jobs / Pop	High Inc. Jobs / Pop	Commute Distance
I. DiD										
New*post	8.98	62.82	31.66	-11.03	42.19	0.02	0.01	-0.00	0.02	-0.30***
	[40.86]	[109.93]	[53.30]	[37.16]	[51.30]	[0.05]	[0.02]	[0.02]	[0.02]	[0.08]
Existing*post	31.60	100.88	1.54	-20.74	120.08	-0.06	-0.03	-0.05**	0.02	-0.25***
	[29.64]	[122.09]	[50.68]	[31.20]	[79.12]	[0.10]	[0.02]	[0.02]	[0.07]	[0.07]
constant	3562.43***	2082.47***	436.68***	790.12***	855.66*	1.21**	0.18***	0.34***	0.69	16.80***
	[209.11]	[541.12]	[166.08]	[127.20]	[476.12]	[0.51]	[0.06]	[0.07]	[0.49]	[0.51]
II. IV										
New*post	-89.04	242.26	-137.03	-2.58	381.88	-0.07	-0.13	-0.11	0.17	-2.20***
	[141.56]	[449.50]	[150.51]	[125.75]	[303.37]	[0.40]	[0.08]	[0.10]	[0.30]	[0.35]
Existing*post	25.82	110.89	-7.86	-20.27	139.02	-0.06	-0.04	-0.05*	0.03	-0.36***
	[31.60]	[136.75]	[49.31]	[36.14]	[92.46]	[0.12]	[0.02]	[0.03]	[0.09]	[0.07]
First Stage F Stat	33.30	31.38	31.38	31.38	31.38	31.38	31.38	31.38	31.38	31.38
Endogeneity Test (p-value)	0.401	0.467	0.892	0.512	0.444	0.726	0.069	0.658	0.304	0.000
Overidentification Test (p-value)	0.415	0.311	0.165	0.942	0.116	0.691	0.515	0.351	0.395	0.418
Observations	3458	3458	3458	3458	3458	3458	3458	3458	3458	3458

See notes to Table 3.4. Table 3.5 follows the same specifications as the regressions in Table 3.4, except restricting the control group to tracts where *Historical* = 1. Robust standard errors clustered by tracts. *, **, and *** denote 0.1, 0.05, and 0.01 significance levels.

However, the results are more supportive of increased job density when using the continuous proximity to Metro station specification (Table 3.6). As proximity to a Metro station increases, the number of jobs increases although population seems to decrease as well. The IV interpretation is that for every percentage increase in proximity to a Metro station, on average, the total number of jobs in the tract increases by 0.32 percent, population decreases by 0.093 percent, and the total number of jobs per resident increases by 0.095 percent. Additionally, commuting distances also decrease by 0.183 percent for every one percent increase in proximity to Metro stations. Thus, the continuous specification provides stronger evidence of increased jobs in areas closer to Metro stations, possibly from increased economic activity. Specifications using continuous distance may be more appropriate for measuring the effect on job densities – as proximity to Metro stations may influence job growth beyond a 0.8-kilometer walking distance buffer.

While the specification using continuous distance is consistent with the hypothesis that job growth within a tract is driving increased LFP and employment for residents, there is a contradiction with the ACS results. First, the LODES data suggests that most increased jobs is are in high income jobs (Table 3.6). However, recall in Table 3.1 and 3.2 that average earnings were decreasing for residents – although keep in mind this was only statistically significant in the binary treatment variable specification and not in Table 3.3 or 3.4 with the analogous continuous distance measure. Thus, the earnings effect is unclear. Graham et al. (2014) lists several reasons why the ACS and LODES data might not align and may differ substantially for some queries. In terms of earnings, ACS asks the respondent the total earnings for all jobs over the past 12 months, while LODES is a snapshot of one job at April 1st and requires a person to have positive earnings from January through June of the current year to appear in the LODES. A more flexible hourly wage measure will better capture the income distributional effects of public transit infrastructure, which does not exist in either LODES or the ACS.[82]

Additionally, there are broader issues raised by Graham et al. (2014) that questions the reliability of relying using LODES data, or at least in conjunction with ACS, to address job density. For example, LODES data reports place of work based on federally administrative data rather than a

[82] The fixed intervals for classifying low income, medium income, and high-income jobs in LODES might not be appropriate as they are not adjusted to account for any changes in the earnings distribution in California. For example, consider minimum wage policy in California. A worker earning minimum wage in California in 2016 working full-time will equate to approximately $1,733 monthly earnings, well above the low-income threshold. Thus, the effect on low-income jobs may be understated as minimum wage policy changes push workers out of the low-income jobs category although this should have less of an effect on high income jobs.

physical address reported by the individual surveyed. In the case of the former, for some workers, their primary workplace location may not be the same address as the administrative location, particularly for lower-income jobs. Some additional issues are the different methods of handling small sample sizes – LODES employ various noise infusion methods and small sample imputation methods while ACS typically suppresses output. Overall, due to different methods used in the data generating process, I interpret the results from Tables 3.4-3.6 with a degree of caution.

To summarize, there is clear and convincing evidence that Metro stations lead to increased labor force participation and employment, which is robust to a variety of specifications specifying proximity differently or restricting to a feasible subset of tracts as the control group. However, it appears for public transit usage, existing workers reliant on public transportation are substituting buses for more efficient methods in tracts with new stations. Therefore, it is unlikely the employment boost can be explained by expanded usage of the new transit infrastructure. Increased residential employment is more likely a result from increased job densities in areas near transit stations is contributing to greater LFP and employment for residents, although the evidence using the LODES data to measure the effect on job density is less conclusive.

Table 3.6: Effects of Proximity to Metro Stations on Jobs

Data Source Variables	(1) ACS Population	(2) LODES Total Jobs	(3) LODES Low Inc. Jobs	(4) LODES Med Inc. Jobs	(5) LODES High Inc. Jobs	(6) Both Total Jobs / Pop	(7) Both Low Inc. Jobs / Pop	(8) Both Med Inc. Jobs / Pop	(9) Both High Inc. Jobs / Pop	(10) LODES Commute Distance
I. DiD										
KmFromStation	0.012*	-0.074***	-0.045*	-0.072**	-0.163***	-0.020**	-0.008	-0.003	-0.013**	0.011***
	[0.007]	[0.027]	[0.027]	[0.030]	[0.052]	[0.010]	[0.008]	[0.005]	[0.006]	[0.002]
Constant	8.959***	7.558***	6.370***	6.687***	6.005***	0.479***	0.185***	0.239***	0.187***	3.496***
	[0.061]	[0.189]	[0.179]	[0.205]	[0.327]	[0.060]	[0.036]	[0.034]	[0.039]	[0.025]
II. IV										
KmFromStation	0.093**	-0.320***	-0.142*	-0.225**	-0.824***	-0.095***	-0.018	-0.012	-0.063***	0.183***
	[0.038]	[0.085]	[0.085]	[0.092]	[0.134]	[0.025]	[0.014]	[0.013]	[0.015]	[0.018]
III. First Stage Equation										
Hypo Line Distance	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***	0.009***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
First Stage F-Stat	203.76	203.76	203.76	203.76	203.76	203.76	203.76	203.76	203.76	203.76
Endogeneity Test (p-value)	0.011	0.002	0.221	0.077	0.000	0.001	0.397	0.455	0.000	0.000
Observations	4604	4604	4604	4604	4604	4604	4604	4604	4604	4604

See Table 3.4 notes for data description and Table 3.3 notes for DiD/IV description notes. Robust standard errors clustered by tracts. *, **, and *** denote 0.1, 0.05, and 0.01 significance levels.

3.6 Conclusion

Overall, the results indicate increases in labor force participation and employment from proximity to Metro stations, which shows that these stations may be reducing spatial mismatch. The preferred specification, using the binary treatment variable with the restricted subsample of tracts, indicates a 5.49 percentage point increase in the labor force participation rate and 4.93 percentage point increase in employment. I attribute this potentially to increased job density in neighborhoods close to stations, although the evidence using the LODES data is not as conclusive as the ACS. Nevertheless, there is evidence that commuting distances are decreasing for workers and that number of jobs increases the closer a tract is to a Metro station.

Further work on this topic is being conducted with confidential microdata data from the ACS. First, confidential data will allow for multi-period models, pooling together more expansions, including the completion of the Expo line and Gold line extension in 2016. The microdata allows more detailed demographic cuts on earnings for the different subgroups studied. Furthermore, it can provide a robustness check against neighborhood sorting mechanisms by removing out or controlling for individuals who migrated to the tract in the past year. Finally, it can expand more on both job densities and commuting patterns, as it provides detailed information about a person's place of residence and work without sacrificing detailed demographic information (like the LODES data), which could provide a more complete spatial mismatch story.[83] Nevertheless, the paper contributes to a sparse literature on the effects of public transit infrastructure on employment outcomes and distinguishes itself by addressing route placement endogeneity.

[83] For example, LODES only begins to report sex, race, ethnicity, and educational attainment breakdowns starting in 2009.

References

Anderson, M. L. (2014). Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion. *American Economic Review*, 104, 2763–2796.

Angrist, J.D. and Pischke, J. (2009). *Mostly Harmless Econometrics*. Princeton: Princeton University Press

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat. *Journal of Experimental Social Psychology*, 35(1), 29-46. DOI: 10.1006/jesp.1998.1371

Åslund, O., Blind, I., & Dahlberg, M. (2017). All Abroad? Commuter Train Access and Labor Market Outcomes. *Regional Science and Urban Economics*, 67(C), 90-107.

Baade, R. & Matheson, V.A. (2002). Bidding for the Olympics: Fool's Gold? In C.P.B. Editor, M.I. Editor, & S.S. Editor, *Transatlantic Sport: The Comparative Economics of North American and European Sports* (127–51). London: Edward Elgar

Baade, R.A. and Matheson, V.A. (2016). Going for the Gold: The Economics of the Olympics. *Journal of Economic Perspectives*, 30(2), 201-218.

Baumann, R., Engelhardt, B., & Matheson, V.A. (2012). Employment Effects of the 2002 Winter Olympics in Salt Lake City, Utah. *Journal of Economics and Statistics*, 232(3), 308–17

Baum-Snow, N. & Kahn, M.E. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics*, 77(2), 241-263.

Baum-Snow, N. (2007). Did Highway Cause Suburbanization. *Quarterly Journal of Economics*, 122(2), 775-805.

Becker, G. (1957). The Economics of Discrimination. Chicago: University of Chicago Press. ISBN 9780226041049.

Berger, M.C. (1984). Cohort Size and the Earnings Growth of Young Workers. *Industrial and Labor Relations Review,* 37(4), 582-91.

Berlin, N. and Dargnies, M.P. (2016). Gender Differences in Reactions to Feedback and Willingness to Compete. *Journal of Economic Behavior & Organization*, 130, 320-336. DOI: j.jebo.2016.08.002.

Black, S.E., & Strahan, P.E. (2001). The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry. *American Economic Review*, 91(4), 814-831. DOI: 10.1257/aer.91.4.814

Bowes, D.R. & Ihlanfeldt, K.R. (2001). Identifying the Impacts of Rail Transit Stations on Residential Property Values. *Journal of Urban Economics*, 50(1), 1-25

Brooks, L. & Lutz, B. (2019). Vestiges of transit: Urban persistence at a microscale. *Review of Economics and Statistics*, 101(3), 385-399.

Bruckner, J. & Zenou, Y. (2003). Space and Unemployment: The Labor-Market Effects of Spatial Mismatch. *Journal of Labor Economics*, 21(1), 242-262.

Bureau of Labor Statistics, U.S. Department of Labor, The Economics Daily, Women's earnings 83 percent of men's, but vary by occupation on the Internet at https://www.bls.gov/opub/ted/2016/womens-earnings-83-percent-of-mens-but-vary-by-occupation.htm (visited March 16, 2017).

Burtless, G. (1986). Social Security, Unanticipated Benefit Increases, and the Timing of Retirement. *Review of Economic Studies*, 53(5), 781-805.

Card, D. & Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *Quarterly Journal of Economics*, 116(2), 705-746.

Cadinu, M., Rosabianca A., & Kiesner, J., (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, 16(7), 572-578. DOI: 10.1111/j.0956-7976.2005.01577.x

Charles, C., Gneezy, U., & Halladay, B., (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131(A), 141-150. DOI: 10.1016/j.jebo.2016.08.010

Coates, D., & Humphreys, B.R. (2008). Do Economists Reach a Conclusion on Subsidies for Sports Franchises, Stadiums, and Mega-Events? *Econ Journal Watch*, 5(3), 294–315.

Cohen, J.D. (2006). The Aging Nurse Workforce: How to Retain Experienced Nurses. *Journal of Healthcare Management*, 51(4), 233-45.

Colby, S. L., & Ortman, J. M. (2014). The baby boom cohort in the United States: 2012 to 2060 (Current Population Reports P25-1141). Washington, DC: U.S. Census Bureau.

Collegeboard Homepage. (n.d.). Retrieved March 16, 2017, from https://www.collegeboard.org/

Coulson, N.E., Laing, D., & Wang, P. (2001). Spatial Mismatch in Search Equilibrium. *Journal of Labor Economics*, 19(4), 949-972.

Credit, K. (2017). Transit-oriented economic development: The impact of light rail on new business starts in the Phoenix, AZ Region, USA. *Urban Studies* 55(13): 2838-2862.

Desert, M., Preaux, M., & Jund, R., (2009). So Young and Already Victims of Stereotype Threat: Socioeconomic Status and Performance of 6 to 9 Years Old Children on Raven's Progressive Matrices. *European Journal of Psychology of Education*, 24(2), 207-218. DOI: 10.1007/BF03173012

Duranton, G. & Turner, M.A. (2012). Urban Growth and Transportation. *Review of Economic Studies*, 79, 1407-1440.

Dziauddin, M.F. (2019). Estimating Land Value Uplift Around Light Rail Transit Stations in Greater Kuala Lumpur: An Empirical Study Based on Geographically Weighted Regression (GWR). *Research in Transportation Economics,* 74, 10-20.

Elkin, E.N. (2014). *Railtown: The Fight for the Los Angeles Metro Rail and the Future of the City*. Berkeley: University of California Press; First Edition.

El-Geneidy, Ahmed, Michael Grimsrud, Rania Wasfi, Paul Tétreault, & Julien Surprenant-Legault. 2014. "New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas." Transportation 41: 193–210.

Evers, M., de Mooij, R., & van Vuuren, D. (2008). The Wage Elasticity of Labor Supply: A Synthesis of Empirical Estimates. *De Economist*, 156(1), 25-43.

Flood, S., King, M., Ruggles, S. & Warren, J.R. (2017). *Integrated Public Use Microdata Series, Current Population Survey: Version 5.0* [dataset]. Minneapolis, MN: University of Minnesota.

French, E. & Jones, J. (2012). Public Pensions and Labor Supply Over the Life Cycle. *International Tax and Public Finance,* 19(2), 268-87.

Friedberg, L. (2000). The labor supply effects of the social security earnings test. *Review of Economics and Statistics*, 82(1), 429–450.

Fryer, R.G., Levitt, S.D., & List, J.A., (2008). Exploring the Impact of Financial Incentives on Stereotype Threat: Evidence from a Pilot Study. *American Economic Review*, 98(2), 370-375. DOI: 10.1257/aer.98.2.370

Ganley, C.M., Mingle, L.A., Ryan, A.M., Ryan, K., Vasilyeva, M., & Perry, M., (2013). An examination of stereotype threat effects on girls' mathematics performance. *Educational Psychology*, 49(10), 1886-1897. DOI: 10.1037/a0031412

Gneezy, U., Niederle, M., & Rustichini, A., (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, 118(3), 1049-1074. DOI: 10.1162/00335530360698496

Gobillon, L., Selod, H., & Zenou, Y (2007). "The Mechanisms of Spatial Mismatch". *Urban Studies*.

Graham, M.R., Kutzbach, M.J., & McKenzie, B. Design Comparison Between LODES and ACS Commuting Data Products. Working Paper CES-WP-14-38, Center for Economic Studies, U.S. Census Bureau, 2014.

Grove, R. D., & Hetzel, A. M. (1968). *Vital statistics rates in the United States 1940–1960* (Report). Washington, DC: National Center for Health Statistics, U. S. Government Printing Office.

Gruber, J., & Wise, D. A. (2007). *Social security programs and retirement around the world*. Chicago, IL: University of Chicago Press.

Gustman, A. L., & Steinmeier, T. L. (2005). The social security early entitlement age in a structural model of retirement and wealth. *Journal of Public Economics*, 89(2-3), 441–463.

Healy, A. and Pate, J. (2011), Can Teams Help to Close the Gender Competition Gap?. *The Economic Journal*, 121: 1192-1204.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 1, pp. 307–458). Amsterdam, The Netherlands: Elsevier.

Heilmann, K. (2018). Transit access and neighborhood segregation. Evidence from the Dallas light rail system. *Regional Science and Urban Economics* 73(C), 237-250.

Hellerstein, J.K., Neumark, D., & McInerney, M. (2008). Spatial Mismatch or Racial Mismatch? *Journal of Urban Economics*, 64(2): 464-479.

Hellerstein, J.K., Neumark, D., & Troske, K., (1999). Wages, Productivity, and Worker Characteristics. *Journal of Labor Economics*, 17(3), 409-446.

Hoff, K., & Pandey, P., (2006). Discrimination, Social Identity, and Durable Inequalities. *American Economic Review*, 96(2), 206-211. DOI: 10.1257/000282806777212611

Holzer, H.J. (1991). The Spatial Mismatch Hypothesis: What Has the Evidence Shown? *Urban Studies*, 28(1).

Holzer, H.J. & Reaser, J. (2000). Black Applicants, Black Employees, and Urban Labor Market Policy. *Journal of Urban Economics*, 48(3): 365-387.

Holzer, H.J., Quigley, J.M. & Raphael, S. (2003). Public Transit and the Spatial Distribution of Minority Employment: Evidence from a Natural Experiment. *Journal of Policy Analysis and Management*, 22(3), 415-441.

Hotchkiss, J., Moore, R., & Zobay, S.M. (2003). Impact of the 1996 Summer Olympic Games on Employment and Wages in Georgia. *Southern Economic Journal* 69(3): 691–704.

https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2016.pdf, Accessed September 15, 2017.

Hu, R. (2017). The Impact of Rail Transit on the Distribution of New Housing Projects in Beijing. *Urban Studies*, 54(8), 1867–86.

Ihlanfeldt, K.R. & Sjoquist, D.L. (1990). Job Accessibility and Racial Differences in Youth Employment Rates. *The American Economic Review*, 80(1): 267-276.

Ihlanfeldt, K.R. & Sjoquist. (1991). The Effect of Job Access on Black and White Youth Employment: A Cross-Sectional Analysis. *Urban Studies*, 28(2): 255-265.

Iriberri, N., & Rey-Biel, P., (2017). Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Journal of Economic Behavior & Organization*, 135(C), 99-111.

Johnson, R. W., Kawachi, J., & Lewis, E. K. (2009). *Older workers on the move: Recareering in later life* (AARP Public Policy Institute, Research Report #2009-08). Washington, DC: AARP.

Kahn, M.E. (2007). Gentrification Trends in New Transit-Oriented Communities: Evidence from 14 Cities That Expanded and Built Rail Transit Systems. *Real Estate Economics*, 35, 155-182.

Kain, J.F. (1968). Housing Segregation, Negro Employment, and Metropolitan Decentralization. *Quarterly Journal of Economics*, 82, 175-197.

Kain, J.F. (1992). The Spatial Mismatch Hypothesis: Three Decades Later. *Housing Policy Debate*, 3, 371-460.

Kamas, L. & Preston, A., (2012). The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, 83(1), 82-97.

Kaseem, D. (2018). Does Electrification Cause Industrial Development? Grid Expansion and Firm Turnover in Indonesia. Working Paper.

Kawabata, M. (2003). Job Access and Employment among Low-Skilled Workers in US Metropolitan Areas. *Environment and Planning A: Economy and Space*, 35(9), 1651-1668.

Korenman, S., & Neumark, D. (2000). Cohort crowding and youth labor markets: A cross-national analysis. In D. Blanchflower & R. Freeman (Eds.), *Youth employment and joblessness in advanced countries* (pp. 57–106). Chicago, IL: University of Chicago Press.

Kaseem, Dana. "Does Electrification Cause Industrial Development? Grid Expansion and Firm Turnover in Indonesia." Job Market Paper.

Kawabata, Mizuki. (2003). Job Access and Employment among Low-Skilled Workers in US Metropolitan Areas. *Environment and Planning A: Economy and Space*, 35(9): 1651-1668.

L.A. Metro. https://www.metro.net/. Accessed Mar 1, 2018.

Long, H. (2016, September 29). Female CEOs are at record level in 2016, but it's still only 5%. *CNN*. http://money.cnn.com/2016/09/29/investing/female-ceos-record-high/

Linder, F. E., & Grove, R. D. (1947). *Vital statistics rates in the United States 1900–1940* (Report). Washington, DC: U.S. Government Printing Office.

Lipscomb, M., Mobarak, A.M., & Barham, T. (2013). Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. *American Economic Journal: Applied Economics*, 5(2): 200-231.

Logan, J.R., Xu, Z., & Stults, B.J. (2014). Interpolating US Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database. *The Professional Geographer,* 66(3), 412–420.

Logan, J.R., Stults, B.J., & Xu, Z. (2016). Validating Population Estimates for Harmonized Census Tract Data, 2000–2010. *Annals of the American Association of Geographers*. http://www.tandfonline.com/doi/full/10.1080/24694452.2016.1187060.

Los Angeles County GIS Data Portal. *2006 10-foot Digital Elevation Model (DEM) – LARIAC – Public Domain*. Retrieved from: http://egis3.lacounty.gov/dataportal/2011/01/26/2006-10-foot-digital-elevation-model-dem-public-domain/

Macunovich, D. J. (1999). The fortunes of one's birth: Relative cohort size and the youth labor market in the United States. *Journal of Population Economics*, 12, 215–272.

Madden, J.F. (1985). Urban Wage Gradients: Empirical Evidence. *Journal of Urban Economics*, 18, 201-301.

Maestas, N. (2010). Back to work: Expectations and realizations of work after retirement. *Journal of Human Resources*, 45, 718–748.

Marshall, Colin (2016, April 25). Stories of cities #29: Los Angeles and the 'Great American Streetcar Scandal'. *The Guardian*. https://www.theguardian.com/cities/2016/apr/25/story-cities-los-angeles-great-american-streetcar-scandal

Mendelson, Aaron (2015, Feb 19). How Late are Los Angeles Buses and Trains? Depends Which Line You're Riding. *SCPR*. https://www.scpr.org/news/2015/02/19/49937/how-late-are-los-angeles-buses-and-trains-depends/

Mills, Edwin. 1972. *Urban Economics*. Scott Foresman

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517.

Morin, L.-P. (2015). Cohort size and youth earnings: Evidence from a quasi-experiment. *Labour Economics*, 32, 99–111.

Mueller, A. I., Rothstein, J., & von Wachter, T. M. (2016). Unemployment insurance and disability insurance in the Great Recession. *Journal of Labor Economics*, 34, S445–S475.

Munnell, A. H. (2015). *The average retirement age – An update* (Report No. 15-4). Chestnut Hill, MA: Center for Retirement Research at Boston College. Retrieved from http://crr.bc.edu/wp-content/uploads/2015/03/IB_15-4.pdf

National Center for Health Statistics. (1934). *Birth, stillbirth, and infant mortality statistics for the birth registration area of the United States 1931*. Washington, DC: U.S. Government Printing Office.

National Center for Health Statistics. (1934–1938). *Birth, stillbirth, and infant mortality statistics for the continental United States, the Territory of Hawaii, the Virgin Islands 1933 (various years, 1932–1936*). Washington, DC: U.S. Government Printing Office.

National Center for Health Statistics. (1939–2005). *Vital statistics of the United States (various years, 1937–2003*). Washington, DC: U.S. Government Printing Office.

National Center for Health Statistics. (1974–1975). *Summary report final natality statistics (various years,1970–1974*). Rockville, MD: National Center for Health Statistics.

National Center for Health Statistics. (1976–1977). *Advance report final natality statistics (various years, 1974–1975*). Rockville, MD: National Center for Health Statistics.

National Center for Health Statistics. (1978–1980). *Advance report final natality statistics (various years, 1976–1978*). Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. (1982–1996). *Advance report of final natality statistics (various years,1980–1994*). Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. (1984). *Vital statistics of the United States 1979: Volume 1—Natality*. Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. (1997–1998). *Report of final natality statistics (various years, 1995–1996*). Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. (1999–2017). *Births: Final data (various years, 1997–2015).* Hyattsville, MD: National Center for Health Statistics.

Neumark, David, Junfu Zhang, and Stephen Ciccarella. 2008. "The effects of Wal-Mart on local labor markets." Journal of Urban Economics, 63, 405-430.

Niederle, M., Segal, C., & Vesterlund, L. (2013). How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness. *Management Science*, 59(1), 1-16.

Niederle, M., & Vesterlund, L., (2007). Do Women Shy Away From Competition? Do Men Compete Too Much?. *The Quarterly Journal of Economics*, 122(3), 1067-1101. DOI: 10.1162/qjec.122.3.1067

O'Sullivan, S., & Morrall, J. (1996). Walking distances to and from light-rail transit stations. *Transportation Research Record*, 1538, 19–26.

Old-Age Survivors Insurance and Disability Insurance (OASDI). (2016). *The 2016 annual report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. Washington, DC: U.S. Government Printing Office.

Redding, S.J. & Turner, M.A. Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics*, 5, 1339-1398.

Reinhardt, U. (2003). Does the aging of the population really drive the demand for health care? *Health Affairs*, 22(6), 27–39.

Richardson, H. W. (1995). Economies and diseconomies of agglomeration. In H. Giersch (Ed.), *Urban agglomeration and economic growth* (pp. 122–155). Berlin, Germany: Springer.

Rodriguez, C., Gallego, J.M., Martinez, D., Montoya, S., & Peralta-Quiros, T. (2015) Examining Implementation and Labor Market Outcomes of Targeted Transit Subsidies: Subsidy by Sistema Nacional de Selección de Beneficiarios for Urban Poor in Bogotá, Colombia. *Transportation Research Record: Journal of the Transportation Research Board*, 2581(1), 9-17.

Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators. *PloS one*, 11(1)

Phillips, D.C. (2014). Getting to Work: Experimental evidence on job search and transportation costs. *Labour Economics*, 29, 72-82.

Polachek, S.W., (1981). Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure. *Review of Economics & Statistics*, 63(1), 60-69.

Sanchez, T. (1999). The Connection Between Public Transit and Employment. *Journal of the American Planning Association*, 65(3): 284-296.

Sari, F. (2015). Public Transit and Labor Market Outcomes: Analysis of the Connections in the French Agglomeration of Bordeaux. *Transportation Research Part A: Policy and Practice*, 78, 231-251

Schmader, T., Johns, M., & Forbes, C. (2008). An Integrated Process Model of Stereotype Threat Effects on Performance. *Psychological Review*, 115(2), 336–356.

Schuetz, J. (2015). Do Rail Transit Stations Encourage Neighborhood Retail Activity? *Urban Studies*. 52(14), 2699-2723.

Schuetz, J., Giuliano, G., & Shin, E.J. (2018). Can a Car-Centric City Become Transit Oriented? Evidence from Los Angeles. *Cityscape* 20(1), 167-190.

Severen, C. Commuting, labor, and housing market effects of mass transportation: Welfare and identification. Working Paper.

Shepard, R.N., Metzler, J., 1971. Mental rotation of three dimensional objects. *Science*, 171, 701–703.

Shih, Margaret J.; Pittinsky, Todd L.; Ho, Geoffrey C. (2011), "Stereotype boost: positive outcomes from the activation of positive stereotypes", in Inzlicht, Michael; Schmader, Toni, Stereotype threat: theory, process, and application, New York, New York: Oxford University Press, pp. 5–6, 141–143, ISBN 9780199732449.

Shurchkov, O., (2012). Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints. *The Journal of the European Economic Association*, 10(5), 1189-1213.

Spencer, S.J., Steele, C.M., & Quinn, D.M., (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 35(1), 4-28.

Stapleton, D. C., & Young, D. J. (1988). Educational attainment and cohort size. *Journal of Labor Economics*, 6, 330–361.

Steele, C.M., & Aronson, J., (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811.

Steele, C.M., (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613-629.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D.W. Andrews & J. H. Stock (Eds.*), Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (pp. 80–108). Cambridge, UK: Cambridge University Press.

Stoll, M (2005). Geographical Skills Mismatch, Job Search, and Race. *Urban Studies*. 42(4), 695-717.

Stone, J., Lynch, C.I., Sjomeling, M., & Darley, J.M., (1999). Stereotype Threat Effects on Black and White Athletic Performance. *Journal of Personality and Social Psychology*, 77(6), 1213-1227

U.S. Census Bureau. (n.d.). *2017 national population projections tables*. Washington, DC: U.S. Census Bureau. Retrieved from https://census.gov/data/tables/2017/demo/popproj/2017-summary-tables.html

Welch, F. (1979). Effects of cohort size on earnings: The baby boom babies' financial bust. *Journal of Political Economy*, 87, S65–S97.

White, M. (1988). Location Choice and Commuting Behavior in Cities with Decentralized Employment. *Journal of Urban Economics*. 24, 129-152.

Wozniak, D., Harbaugh, W.T., & Mayr U., (2014). The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices. *Journal of Labor Economics*, 32(1), 161-198.

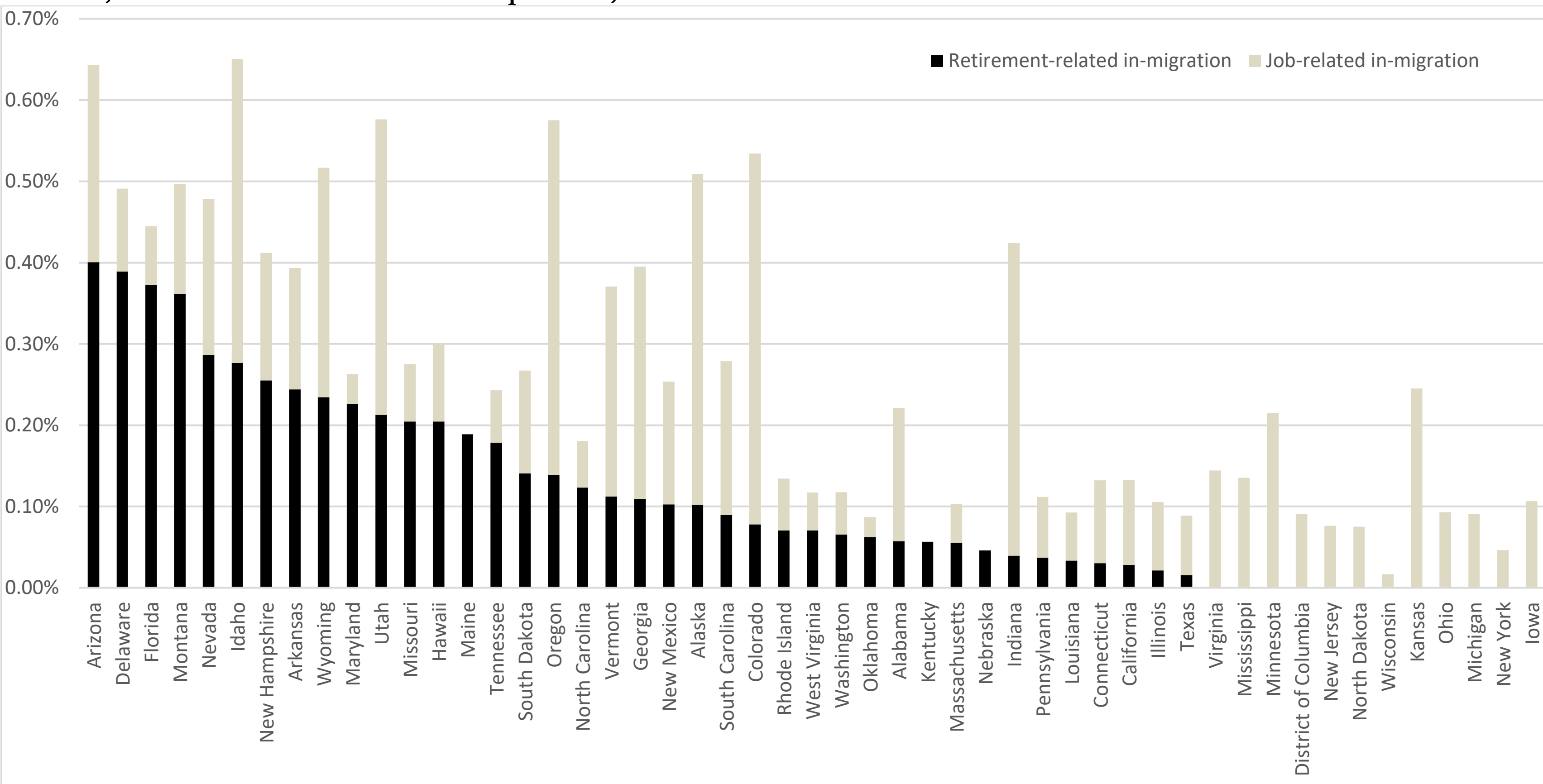
Wright, R. E. (1991). Cohort size and earnings in Great Britain. Journal of Population Economics, 4, 295–305.

Zheng, S. & Kahn, M.E. (2013). China's Bullet Trains Facilitate Market Integration and Mitigate the Cost of Megacity Growth. *Proceedings of the National Academy of Sciences in the United States of America*, 110(14), E1248-E1253.

Zhou, Z., Chen, H., Han, L. and Zhang, A. (2020). The Effect of a Subway on House Prices: Evidence from Shanghai. *Real Estate Economics*.

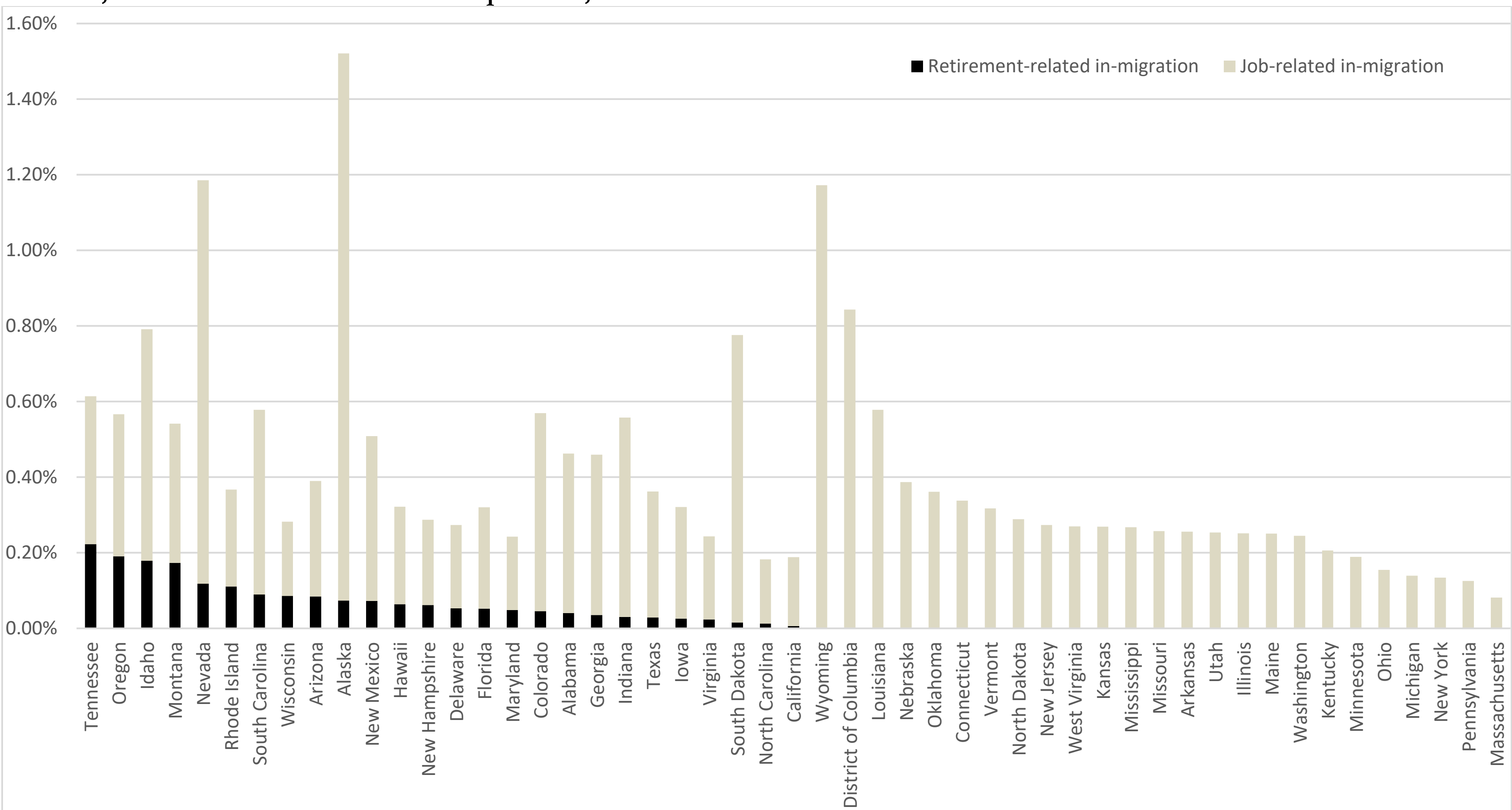
Appendix A: Additional Figures

Appendix Figure A1: One-year Interstate In-Migration of 60-69 Year-Olds for Retirement-Related and Work-Related Reasons, as Percent of 60-69 Year-Old Population, 2008-2016



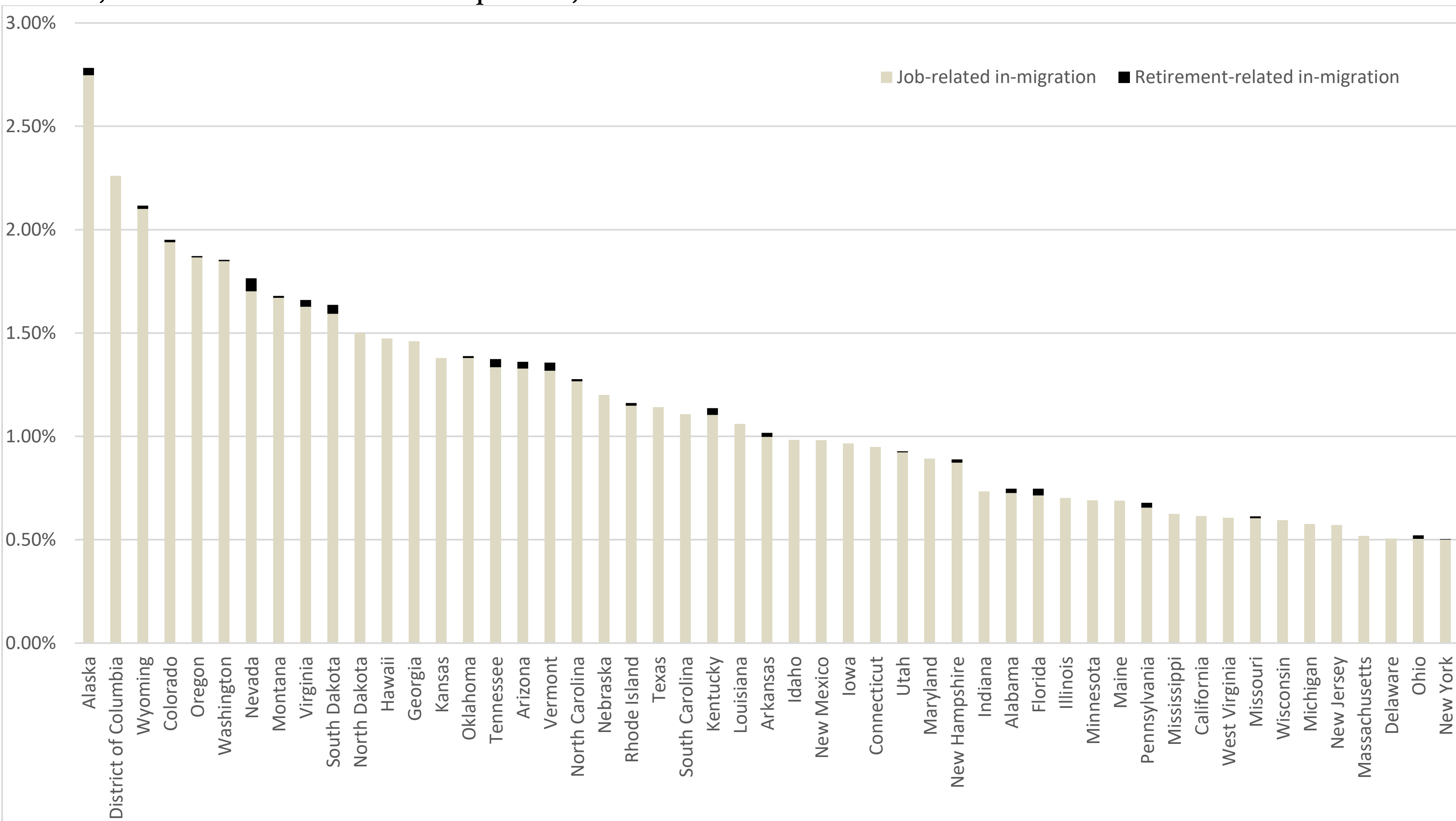
Notes: Based on CPS Annual Social and Economic Supplement (ASEC) data, 2008-2016. Data are constructed using survey weights

Appendix Figure A2: One-year Interstate In-Migration of 50-59 Year-Olds for Retirement-Related and Work-Related Reasons, as Percent of 50-59 Year-Old Population, 2008-2016



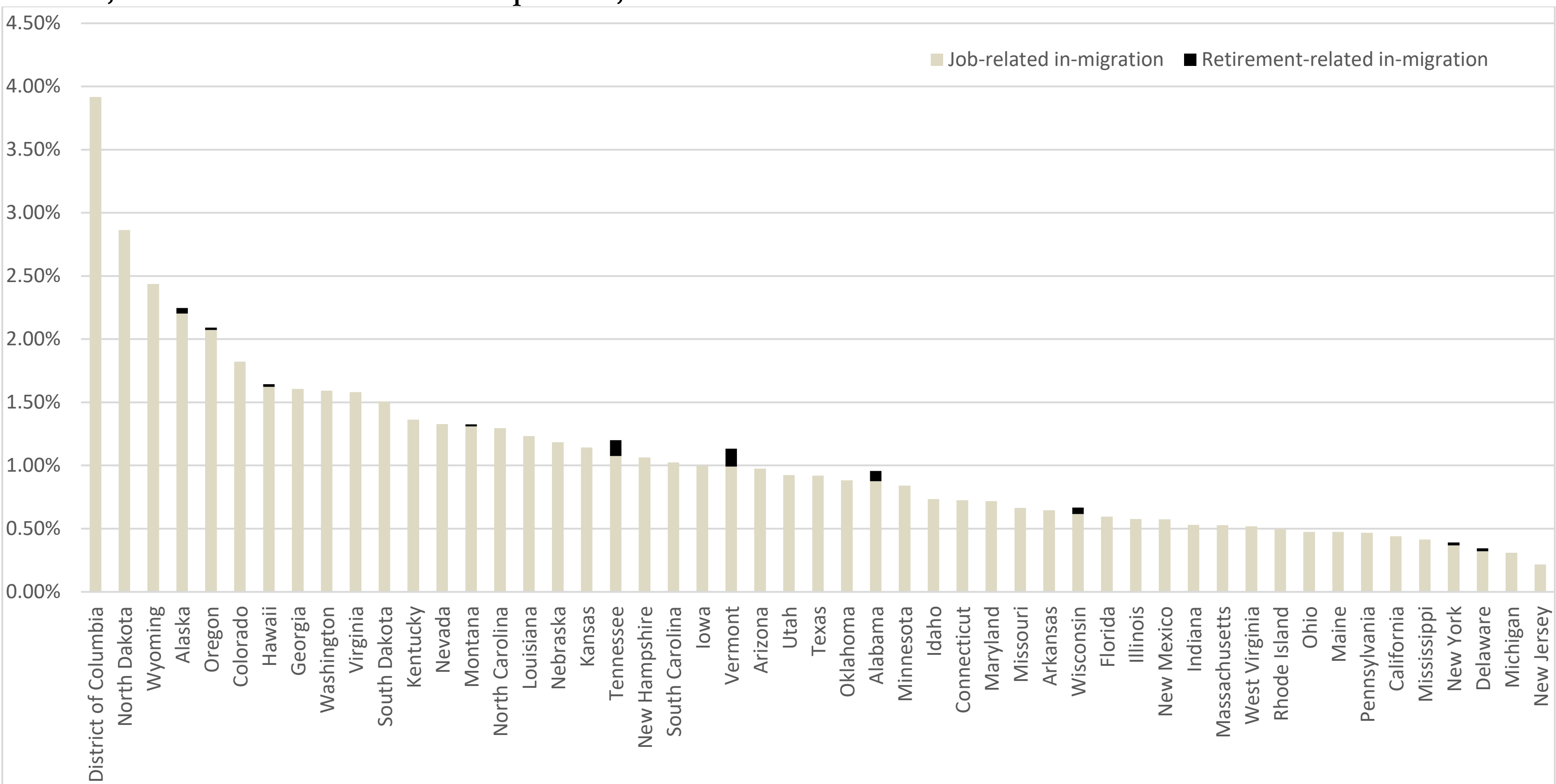
Notes: Based on CPS Annual Social and Economic Supplement (ASEC) data, 2008-2016. Data are constructed using survey weights.

Appendix Figure A3: One-year Interstate In-Migration of 25-49 Year-Olds for Retirement-Related and Work-Related Reasons, as Percent of 25-49 Year-Old Population, 2008-2016



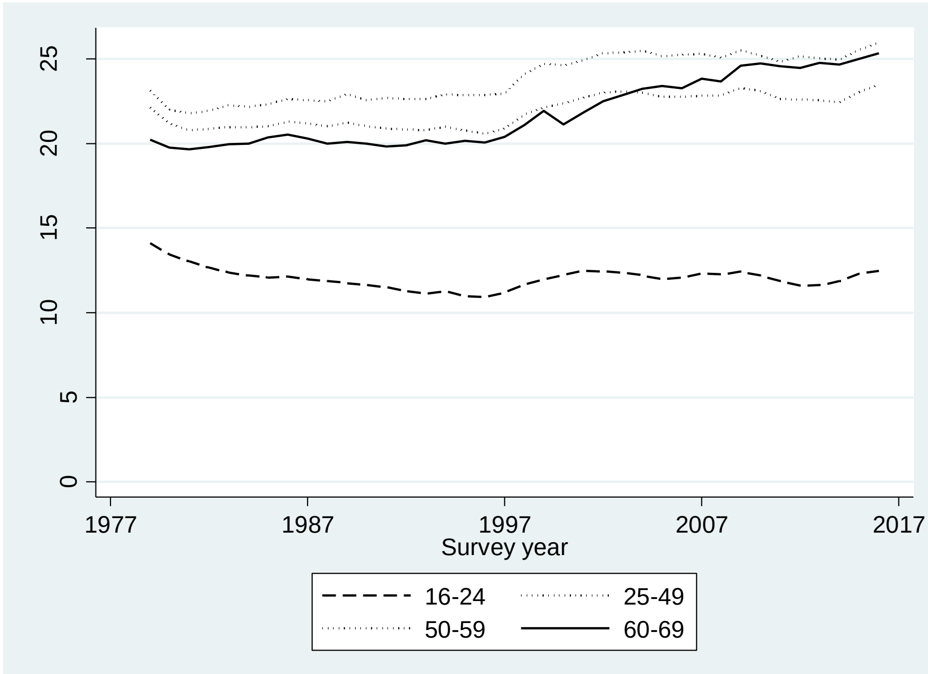
Notes: Based on CPS Annual Social and Economic Supplement (ASEC) data, 2008-2016. Data are constructed using survey weights.

Appendix Figure A4: One-year Interstate In-Migration of 16-24 Year-Olds for Retirement-Related and Work-Related Reasons, as Percent of 16-24 Year-Old Population, 2008-2016



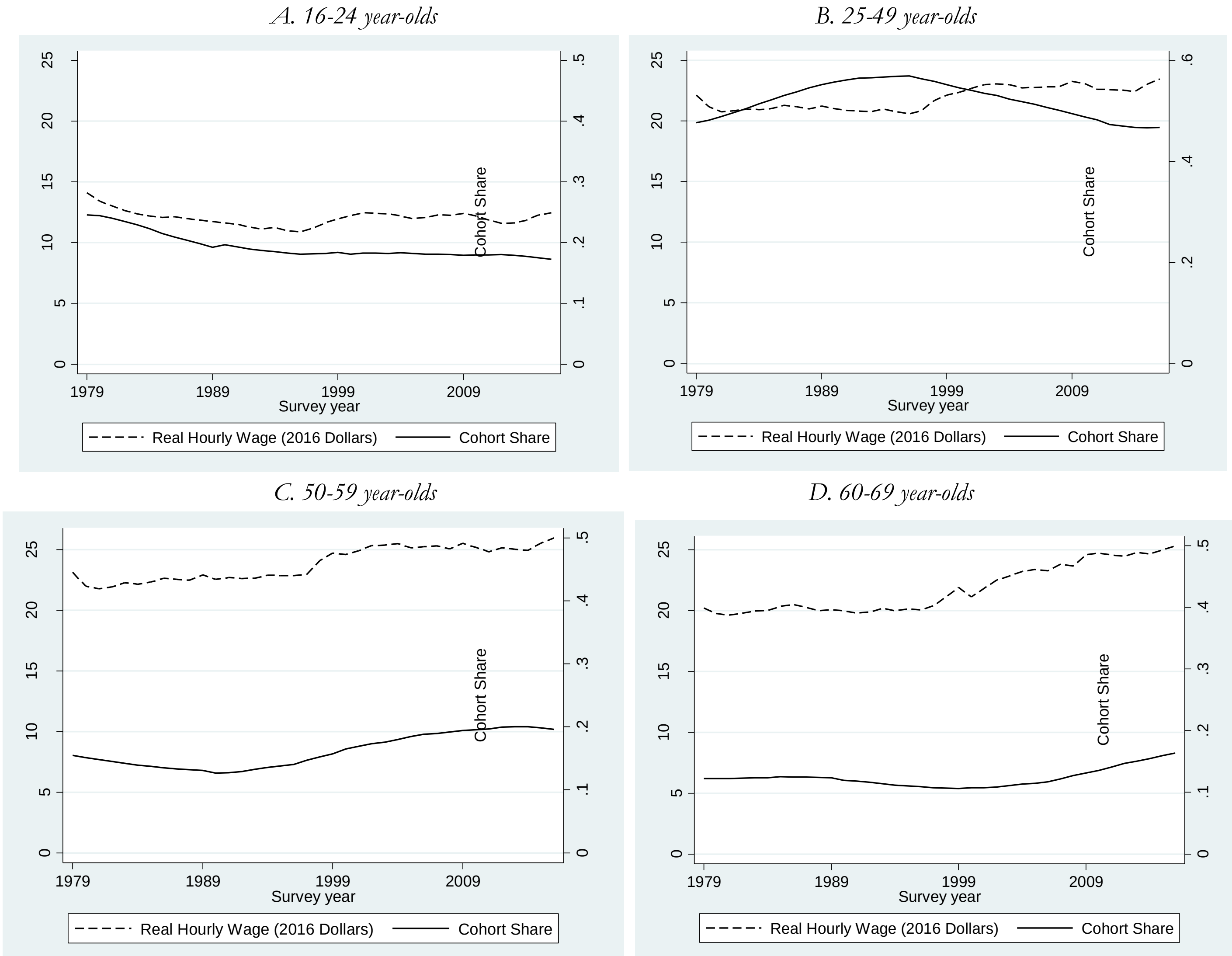
Notes: Based on CPS Annual Social and Economic Supplement (ASEC) data, 2008-2016. Data are constructed using survey weights.

Appendix Figure A5: Real Hourly Wages by Age Group, Over Time



Source: Census Population Survey (CPS) monthly basic files and Outgoing Rotation Group (ORG) files (1979-2016). Hourly wages are used if available or constructed from average weekly earnings divided by usual hours worked if not available. Individuals with wages below half a state's minimum wage or above $200/hour in 2016 dollars are dropped. A state panel on wages is first constructed from the ORG files by aggregated hourly wages for each state and year for each age group. The figure is created from a weighted averages of states' average wages in 2016 dollars, weighted by state population.

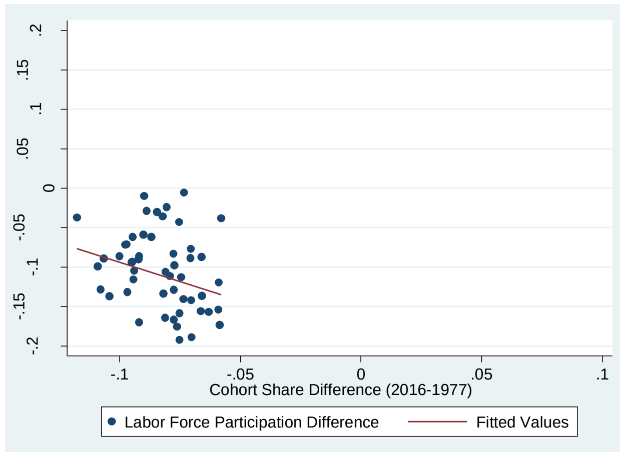
Appendix Figure A6: Wages and Cohort Shares by Age Group, Over Time



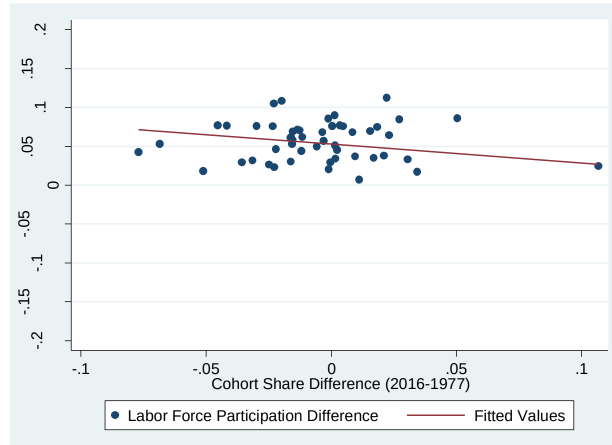
Source: See notes to Figure 2.1 and Appendix Figure A5. Pearson correlation coefficients for 16-24, 25-49, 50-59, and 60-69 year-olds are 0.599, −0.547, 0.894, and 0.627 respectively.

Appendix Figure A7: Labor Force Participation Rates and Cohort Shares by Age Group, 1977-2016 Changes by State

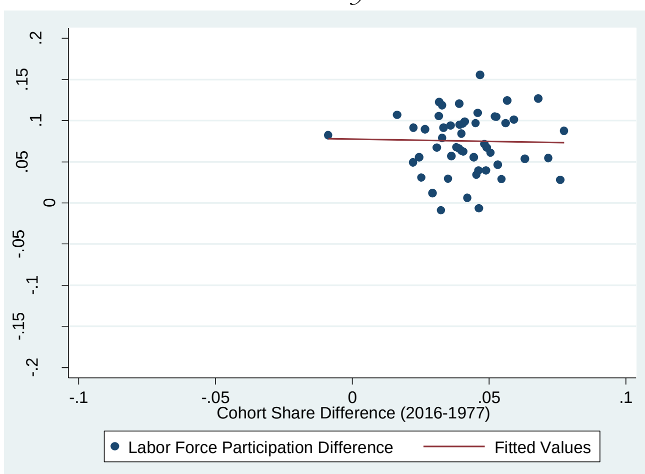
A. 16-24 year-olds



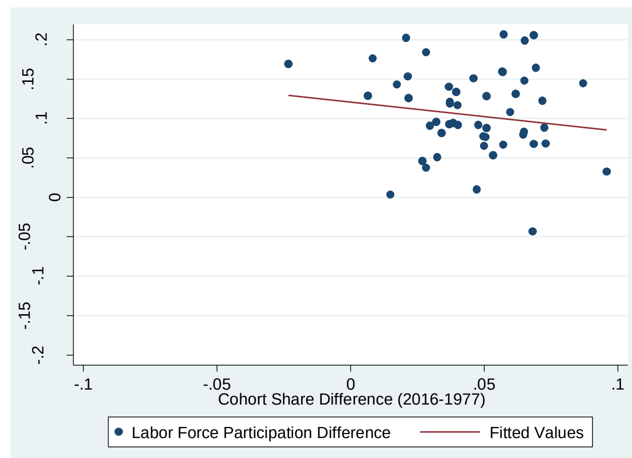
B. 25-49 year-olds



C. 50-59 year-olds



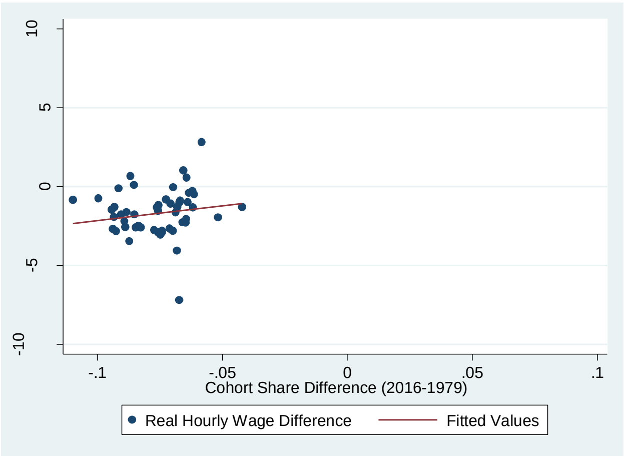
D. 60-69 year-olds



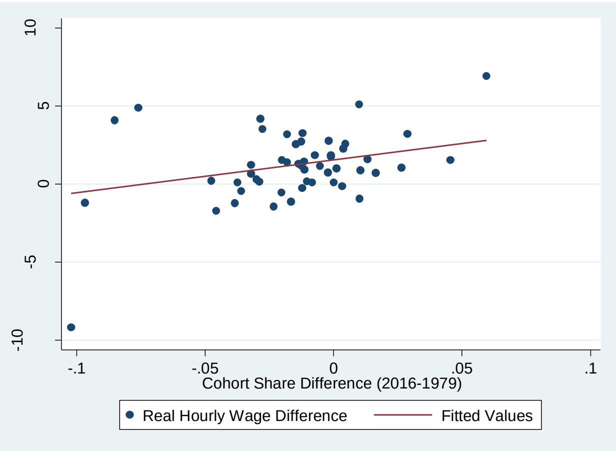
Source: Data source and series construction are explained in notes to Figure 2.1. These figures plot the 1977 to 2016 changes, by state. Fitted values are from regressing the changes in LFP on changes in cohort share differences, weighted by state population through 1977-2016. Regression coefficients and standard errors (given in parentheses) for the 16-24 year-old, 25-49 year-old, 50-59, and 60-69 year-old cohorts are −0.979 (0.490), −0.244 (0.198), −0.052 (0.368), and −0.366 (0.346) respectively.

Appendix Figure A8: Wages and Cohort Shares by Age Group, 1979-2016 Changes by State

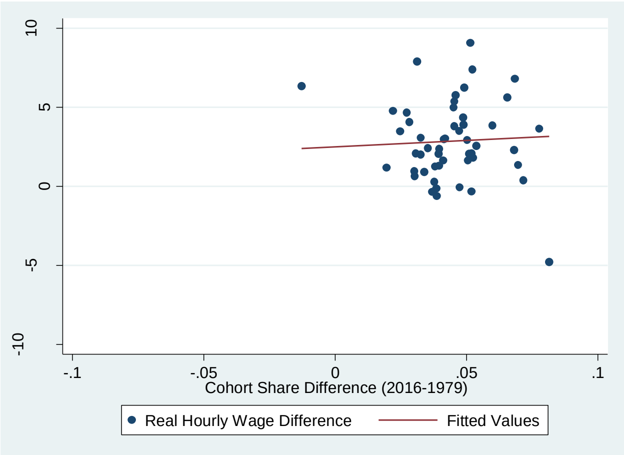
A. 16-24 year-olds



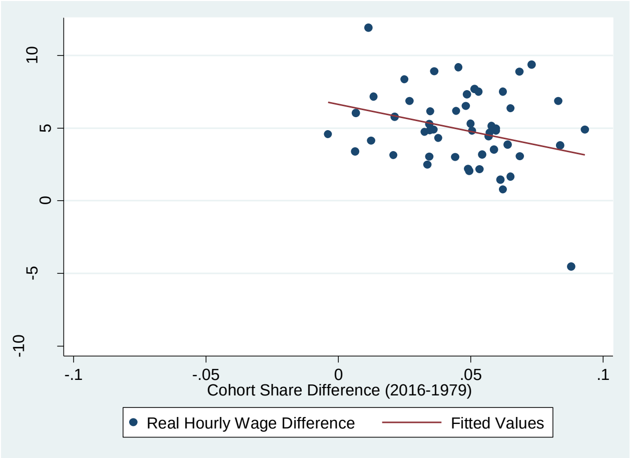
B. 25-49 year-olds



C. 50-59 year-olds



D. 60-69 year-olds



Source: Data source and series construction are explained in Figure 2.1 and Appendix Figure A5. Regression coefficients and standard errors (given in parentheses) for the 16-24 year-old, 25-49 year-old, 50-59 year-old, and 60-69 year-old cohorts are 18.83 (11.65), 21.02 (12.50), 8.18 (27.89), and −37.08 (17.46) respectively. (Note that the implied effect of a one percentage point change in the cohort share is 1/100th of these coefficients.)

Appendix B: Additional Tables

Table B1: Stereotype Threat Language				
Authors	**Field**	**Group**	**Task**	**Stereotype Threat Language**
Steele and Aronson (1995)	Psych	Race	GRE Verbal	Diagnostic of ability
Aronson et al. (1999)	Psych	Race	GRE Math	Packet of info about high performance of Asians in math
Spencer et al. (1999)	Psych	Gender	GRE Math	Test has shown gender differences in the past
Stone et al. (1999)	Psych	Race	Athletic	"Sports intelligence" vs "Natural athletic ability"
Cadinu et al. (2005)	Psych	Gender	Math	"Clear differences in the scores obtained by men and women in logical-mathematical tasks"
Hoff and Pandey (2006)	Econ	Socioeconomic	Mazes	Person's caste revealed (India)
Fryer et al. (2008)	Econ	Gender	SAT Math	"Some academic findings about gender differences in math ability"
Desert et al (2009)	Psych	Socioeconomic	Puzzles	Diagnostic of ability (children)
Irriberri and Rey-Biel (2017)	Econ	Gender	Mental Rotation & Digital Substitution	Own and rival's gender and/or score performance in first round revealed

Table B2: Mental Rotation Score Improvement from Pre-Treatment & Post-Treatment Rounds

Payment Style	(1) PR	(2) PR	(5) Tourn	(6) Tourn
A. Mental Rotation				
Perception Treatment	0.375	0.53	-0.125	0.036
	[1.074]	[1.087]	[1.051]	[1.090]
Update Treatment	-0.5	-0.52	0.938	0.954
	[1.074]	[1.102]	[1.051]	[1.105]
Perception Treatment*Update Treatment	0.75	0.915	-0.563	-0.805
	[1.519]	[1.598]	[1.486]	[1.602]
female	-0.25	-0.525	0.25	0.84
	[1.074]	[1.145]	[1.051]	[1.148]
Perception Treatment*female	-0.5	-0.598	-1.688	-2.088
	[1.519]	[1.533]	[1.486]	[1.537]
Update Treatment*female	0.813	1.504	-0.813	-0.288
	[1.519]	[1.608]	[1.486]	[1.613]
Perception Treatment*Update Treatment*female	-1.5	-2.101	2.375	1.774
	[2.148]	[2.278]	[2.102]	[2.284]
constant	2.125***	1.391	2.125***	3.674
	[0.759]	[3.391]	[0.743]	[3.400]
Covariates Included	No	Yes	No	Yes
N	128	128	128	128
R-Squared	0.022	0.164	0.053	0.151

The dependent variable is the score difference from the post-treatment and pre-treatment rounds (either piece-rate or tournament rounds). Perception and Update treatments are dummy variables that take the value of 1 if the treatment is applied. Included covariates when specified are whether one participates in competitive activity, enjoy competition (1-5), good at competition (1-5), how anxious they felt (1-5), age, whether they are an international student, whether they were a STEM major, and race dummy variables.

Table B3: GRE Verbal Score Improvement from Pre-Treatment & Post-Treatment Rounds

Payment Style	**(1) PR**	**(2) PR**	**(5) Tourn**	**(6) Tourn**
A. Mental Rotation				
Perception Treatment	-0.063	-0.026	0.25	0.349
	[0.575]	[0.603]	[0.567]	[0.591]
Update Treatment	-0.438	-0.625	-0.313	-0.025
	[0.575]	[0.612]	[0.567]	[0.599]
Perception Treatment*Update Treatment	-0.062	0.015	0	-0.492
	[0.813]	[0.886]	[0.802]	[0.868]
female	-0.25	-0.711	-0.063	0.083
	[0.575]	[0.635]	[0.567]	[0.622]
Perception Treatment*female	0.188	0.376	0.125	0.1
	[0.813]	[0.850]	[0.802]	[0.833]
Update Treatment*female	0.125	0.757	0.563	0.412
	[0.813]	[0.892]	[0.802]	[0.874]
Perception Treatment*Update Treatment*female	-0.125	-0.726	-0.313	-0.143
	[1.149]	[1.264]	[1.134]	[1.238]
constant	0.813**	1.947	0.375	0.517
	[0.406]	[1.881]	[0.401]	[1.843]
Covariates Included	No	Yes	No	Yes
N	128	128	128	128
R-Squared	0.022	0.101	0.017	0.109

The dependent variable is the score difference from the post-treatment and pre-treatment rounds (either piece-rate or tournament rounds). Perception and Update treatments are dummy variables that take the value of 1 if the treatment is applied. Included covariates when specified are whether one participates in competitive activity, enjoy competition (1-5), good at competition (1-5), how anxious they felt (1-5), age, whether they are an international student, whether they were a STEM major, and race dummy variables.

Appendix Table B4: Percent Retired within Age Group, 2012-2017

Age Group	Percentage Retired
16-24	0.26%
25-49	0.81%
50-59	6.14%
60-69	40.94%

Source: Census Population Survey (CPS) 2012-2017. The percentage retired is computed from the employment status question, which captures respondents saying that they are not in the labor force due to retirement.

Appendix Table B5: Dates States Joined the Birth Registration System

1915	1916	1917	1919
Connecticut District of Columbia Maine Massachusetts Michigan Minnesota New Hampshire New York Pennsylvania Rhode Island* Vermont	Maryland	Indiana Kansas Kentucky North Carolina Ohio Utah Virginia Washington Wisconsin	California Oregon South Carolina**
1920	**1921**	**1922**	**1924**
Nebraska	Delaware Mississippi New Jersey	Illinois Montana Wyoming	Florida Iowa North Dakota
1925	**1926**	**1927**	**1928**
West Virginia	Arizona Idaho	Alabama Arkansas Louisiana Missouri Tennessee	Colorado Georgia Oklahoma
1929	**1931**	**1932**	**1933**
Nevada New Mexico	Hawaii***	South Dakota	Texas
1945			
Alaska			

Source: National Center for Health Statistics (1947, 1968).
* Rhode Island disappeared from the birth registration for 1919 and 1920.
** South Carolina disappeared from the birth registration for 1925-1927.
*** Hawaii had number of births from individual U.S. Vital Statistics Reports in 1931 as a territorial supplement but does not report crude birth rates until 1950.

Appendix C: Experiment Screenshots

Below contains the following instructions for the experiment and screenshots of the experiment as it is being conducted through zTree. In the rounds where subjects are asked to answer questions, I show the first and last question of that round.

Check In:

Subjects arrive at the Experimental Social Science Laboratory (ESSL) lab and sign in. For this experiment, we provide them a group number that is designed to contain two males and two females in each group. Subjects are not explicitly told the composition of their group but are told to sit in desks next to other members in their group. Students are read a standardized script prior to the experiment for the Experimental Social Science Laboratory (ESSL) that reminds them of their voluntary participation, the duration of the experiment, and rules of conduct before z-tree is initialized.

During the instructions and prior to opening the dividers, I inform subjects
"Some rounds will involve competition within your group where you get paid based on being the highest performer and other rounds you will get paid based on your own performance. Please pay close attention to the payoff structure in each round. Please raise your hand if you are in group 1. Group 2? Group 3? … Group 8?"

I ask subjects to raise their hand as their group number is called so they can see their group members are sitting next to each other.

General Instructions:

Period

1 of 1

Remaining time [sec]: 80

Please read the instructions below

Welcome to this experiment at UC Irvine. Thank you for participating in this study.

You will receive a payment in cash privately at the end of this experiment. The payment will consist of:
- $7 show-up payment
- What you earn during the experiment. Pay attention to each stage's payment structure. Four rounds will be under a piece-rate scheme where you earn a certain amount per question answered correctly and four rounds will be done under a tournament, where the person who has the highest score will win a larger fixed amount (In the event of a tie, this amount will be split equally)

The entire experiment will take place using the computer.

It is important that you do not communicate with any other participants during the experiment.

Please turn off and put away your cell phone.

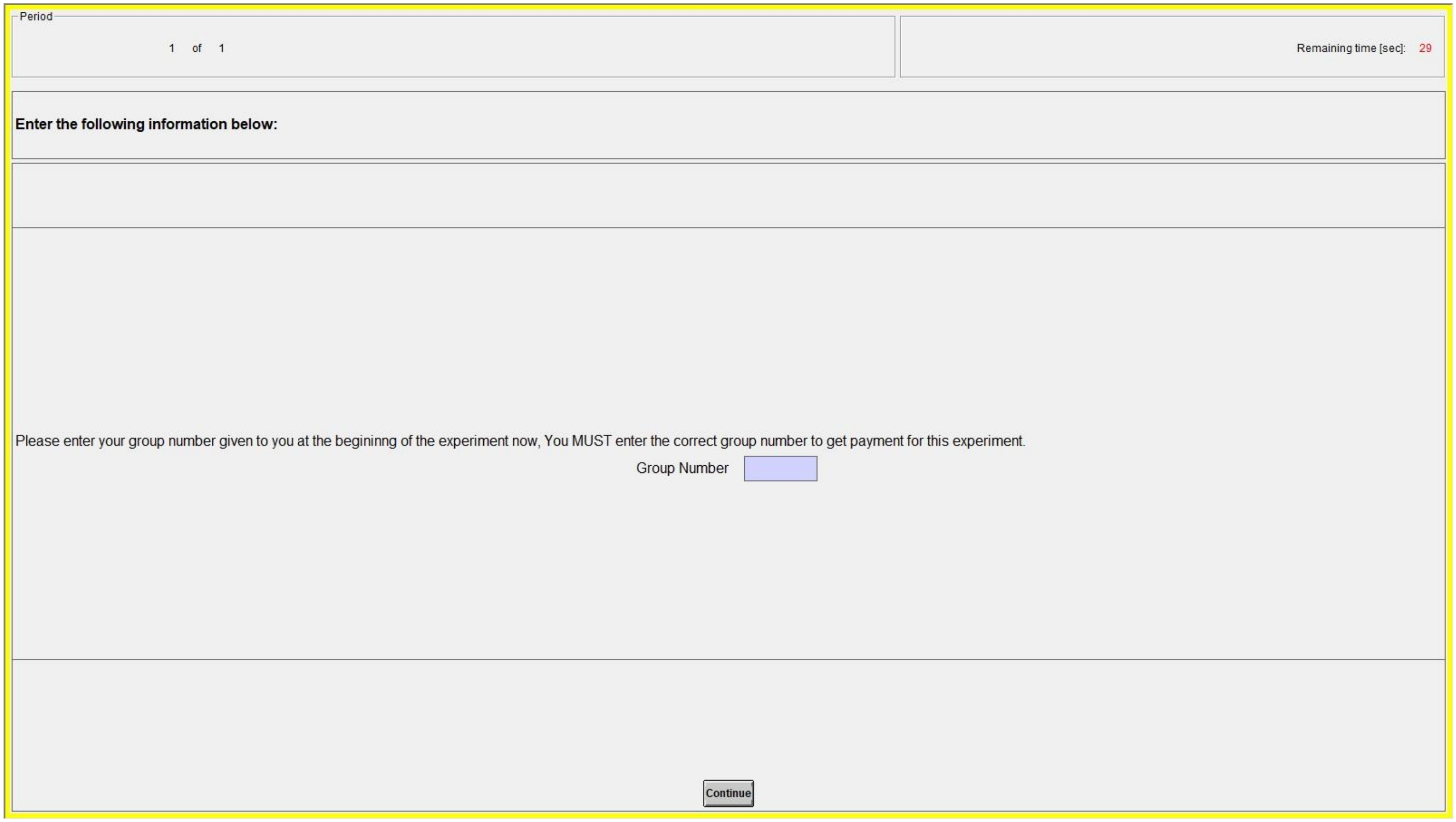
You will be completing four stages of two types of tasks each (a total of eight stages. Each stage will be either 6 minutes 40 seconds or 5 minutes long depending on the task at hand). You will be paid according to your performance in each stage either through a piece-rate (pay dependent on your own performance) or a tournament style (pay dependent on being the highest performer in your assigned group).

The maximum payout in each piece-rate stage is $2 and the maximum payout each tournament is $4 (total of $24 potential variable earnings). You can also earn additional money in some questions in the questionaire.

The timer is in the top right of the screen.

When you are ready, press "Continue" to go to specific instructions for the first stage of the experiment. Note: When you press "Continue," you cannot go back to previous screens.

Continue



Period

1 of 1

Remaining time [sec]: 29

Enter the following information below:

Please enter your group number given to you at the begininng of the experiment now, You MUST enter the correct group number to get payment for this experiment.

Group Number

Continue

Note: I’ve had zero issues with incorrect group numbers being entered.

Mental Rotation Practice Round:

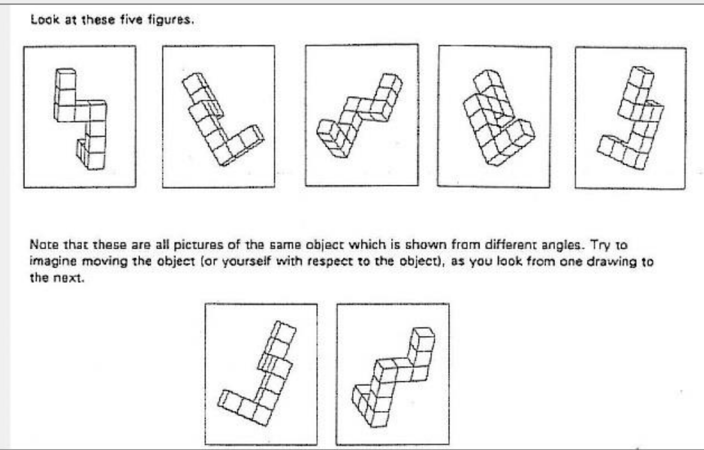
Period

1 of 1

Remaining time [sec]: 60

The mental rotation task requires you to look at photos of 3D figures and decide which figures are identical to a given photo. Take a look at the following example:
When you are finished, click on continue below. You will then be asked to participate in two examples. These will not be scored.
The timer is in the top right of the screen.

Continue



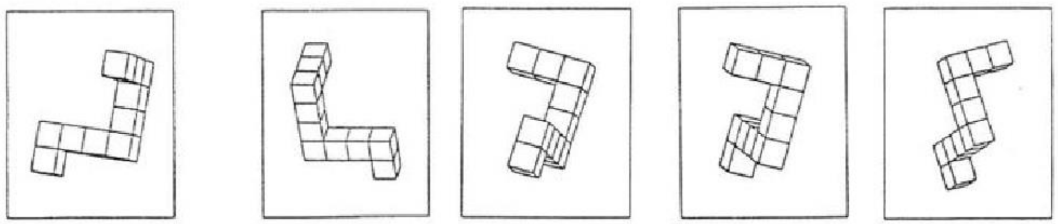
Period

1 of 1

Remaining time [sec]: 27

Example 1:

Choose **two** figures that are identical to the first figure given:



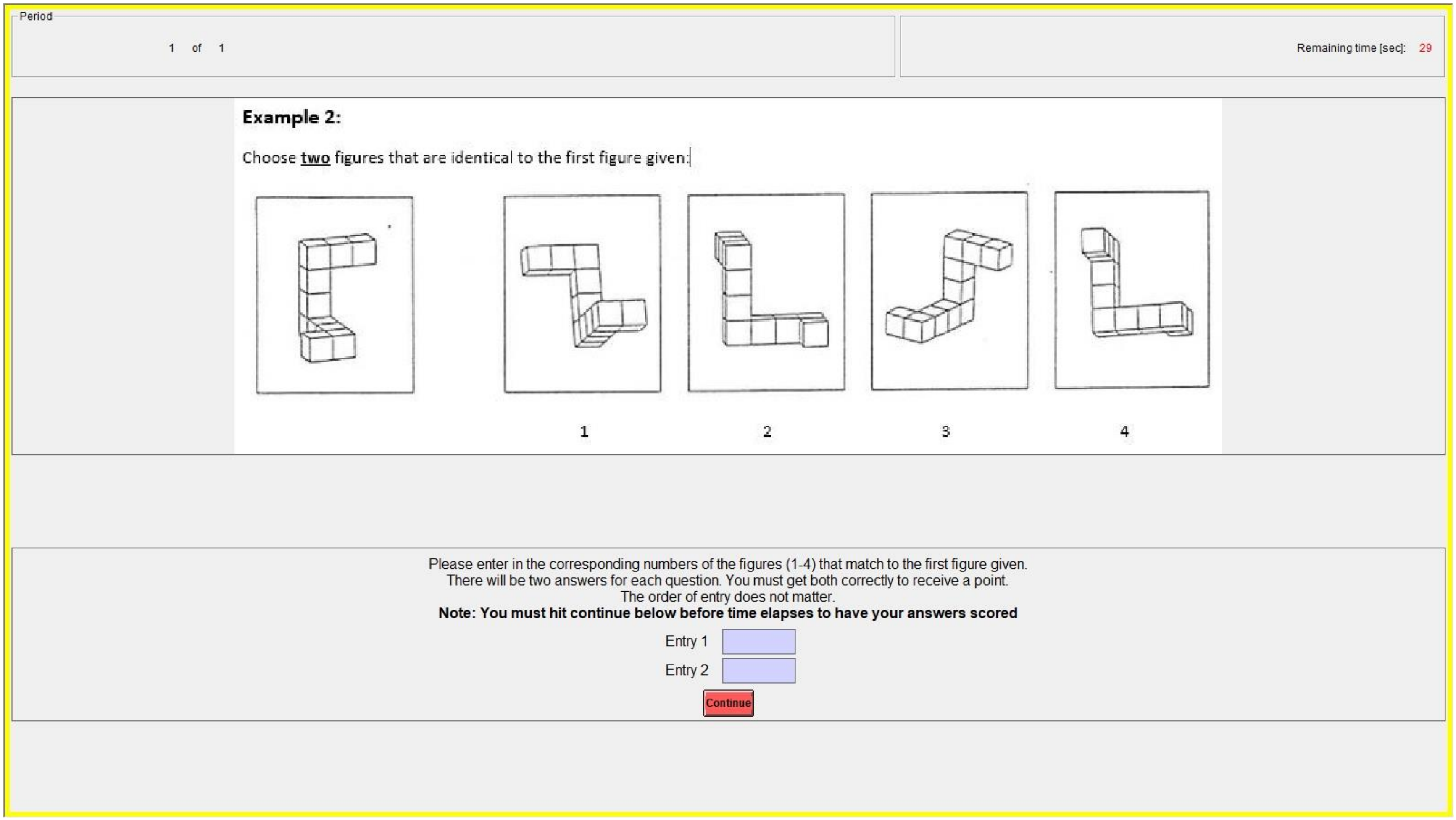
1 2 3 4

Please enter in the corresponding numbers of the figures (1-4) that match to the first figure given.
There will be two answers for each question. You must get both correctly to receive a point.
The order of entry does not matter.
Note: You must hit continue below before time elapses to have your answers scored

Entry 1

Entry 2

Continue

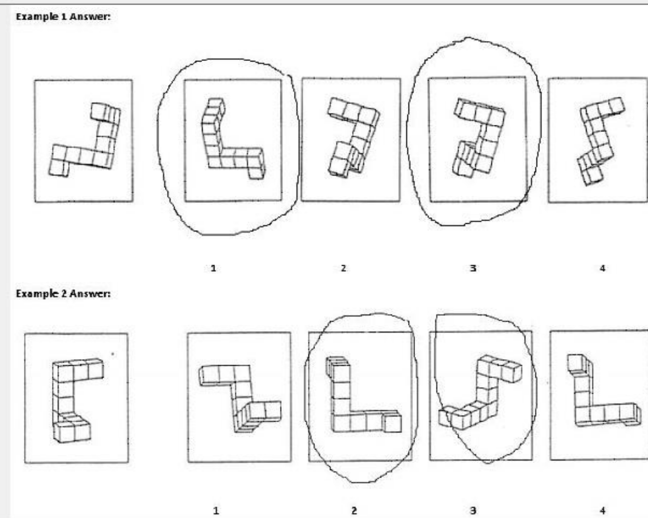


Scores below vary depending on correct answers. Practice round is not formally recorded nor counted towards payoff. The practice rounds are the only time where subjects are given feedback on their performance.

Period

1 of 1

Remaining time [sec]: 14



Remember, you only get **1 point** for choosing both figures correctly.

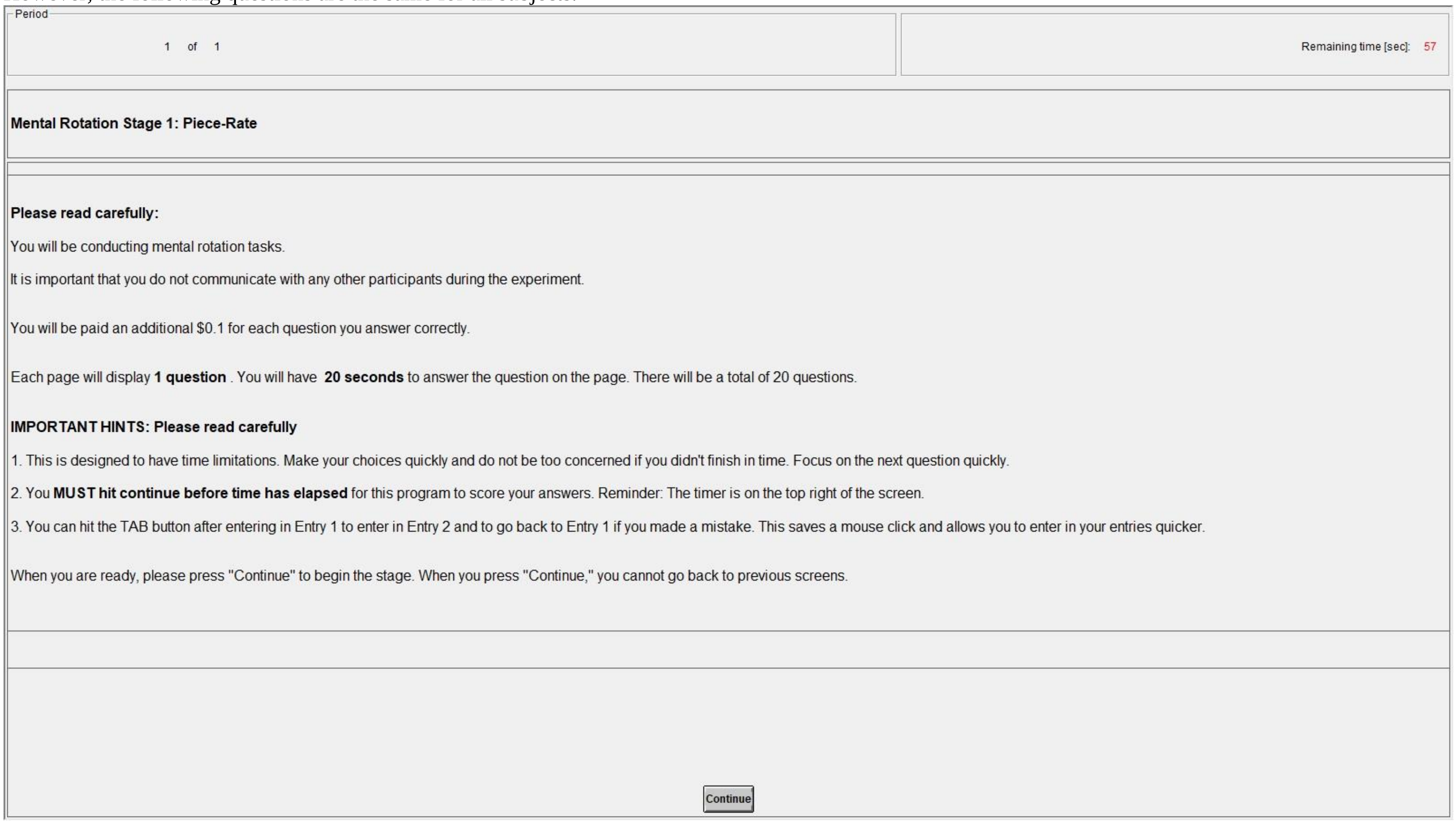
Your Example 1 Score 1

Your Example 2 Score 0

Continue

Mental Rotation Round 1 Instructions:

Note: Odd-numbered groups will begin with the piece-rate payoff structure and even-numbered groups will begin with the tournament pay-off structure. However, the following questions are the same for all subjects.


Period

1 of 1

Remaining time [sec]: 57

Mental Rotation Stage 1: Piece-Rate

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

You will be paid an additional $0.1 for each question you answer correctly.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.

2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.

3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Period

1 of 1

Remaining time [sec]: 60

Mental Rotation Stage 1: Tournament

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.

2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.

3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Mental Rotation Round 1 Questions:

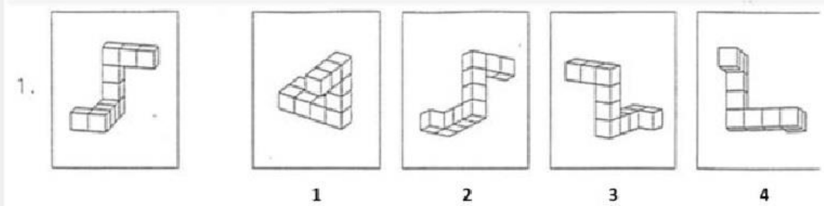
Period

1 of 1

Remaining time [sec]: 19

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 1:



Question 1: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q1 Entry 1:

Q1 Entry 2:

Continue

Period

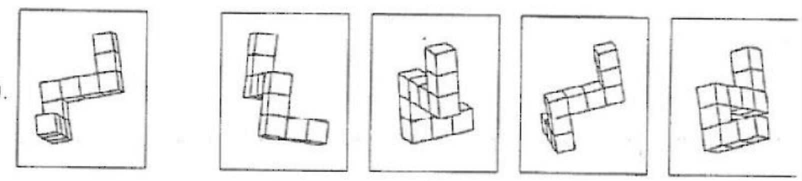
1 of 1

Remaining time [sec]: 20

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 20:

20.



1 **2** **3** **4**

Question 20: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q20 Entry 1:

Q20 Entry 2:

Continue

Period

1 of 1

Remaining time [sec]: 29

Mental Rotation Stage 1 Completion

Congratulations on finishing Mental Rotation Stage 1 of this experiment.

Hit continue when you are ready to move on.

Continue

Mental Rotation Round 2 Instructions:

Note: Odd-numbered groups will continue with the tournament payoff structure and even-numbered groups will continue with the piece-rate payoff structure. However, the following questions are the same for all subjects.

Period

1 of 1

Remaining time [sec]: 60

Mental Rotation Stage 2: Tournament

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.
2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Period

1 of 1

Remaining time [sec]: 59

Mental Rotation Stage 2: Piece-Rate

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

You will be paid an additional $0.1 for each question you answer correctly.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.

2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.

3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Mental Rotation Round 2 Questions:

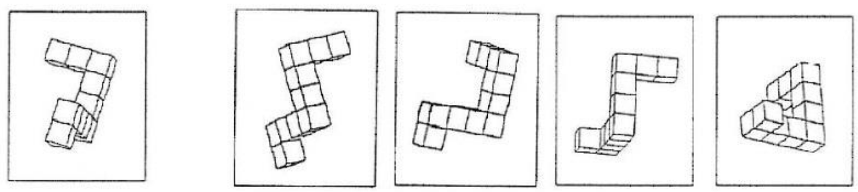
Period

1 of 1

Remaining time [sec]: 20

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 1:



1 2 3 4

Question 1: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q1 Entry 1:

Q1 Entry 2:

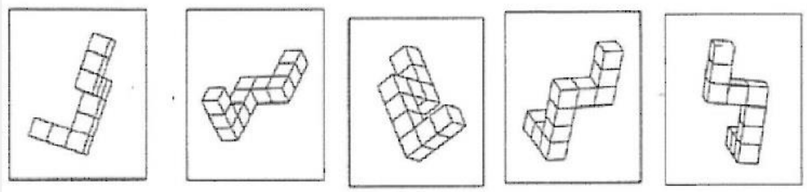
Continue

Period

1 of 1

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 20:



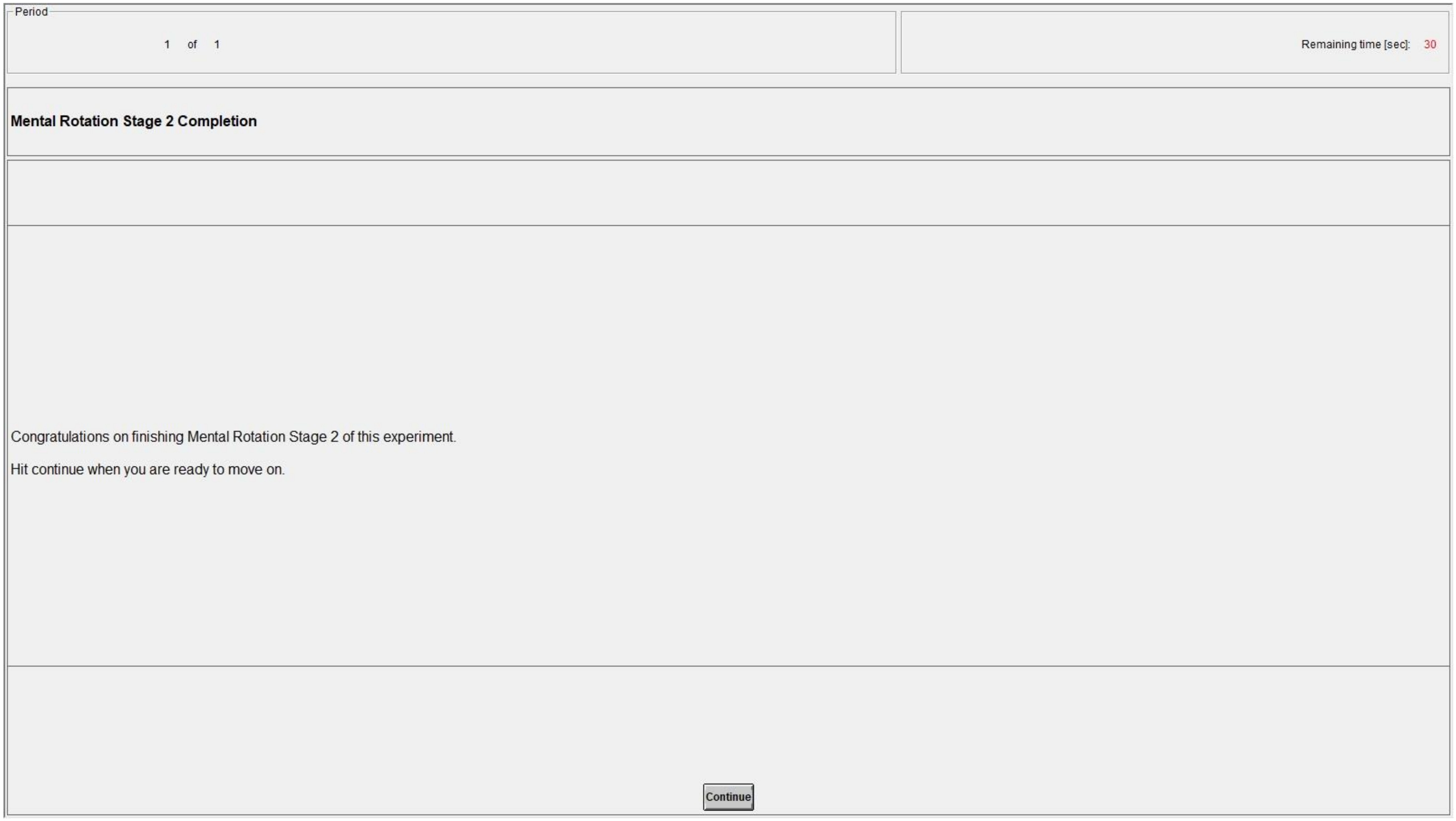
1 2 3 4

Question 20: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q20 Entry 1:

Q20 Entry 2:

Continue


Period
1 of 1
Remaining time [sec]: 30
Mental Rotation Stage 2 Completion
Congratulations on finishing Mental Rotation Stage 2 of this experiment.
Hit continue when you are ready to move on.
Continue

GRE Verbal Practice Round:

Period

1 of 1

Remaining time [sec]: 30

The next two sections will take samples directly from the GRE Verbal Reasoning section.

In particular, the questions will be drawn from sentence completion types of questions. You will see examples of the types of questions on the next page.

When you are finished, click on continue. You will then be asked to participate in two examples. These will not be scored.

Continue

Period

1 of 1

Remaining time [sec]: 60

Many find it strange that her writing is thought to be tortuous; her recent essays, although longer than most of her earlier essays, are extremely __________.

(A) painstaking

(B) tedious

(C) insightful

(D) sophisticated

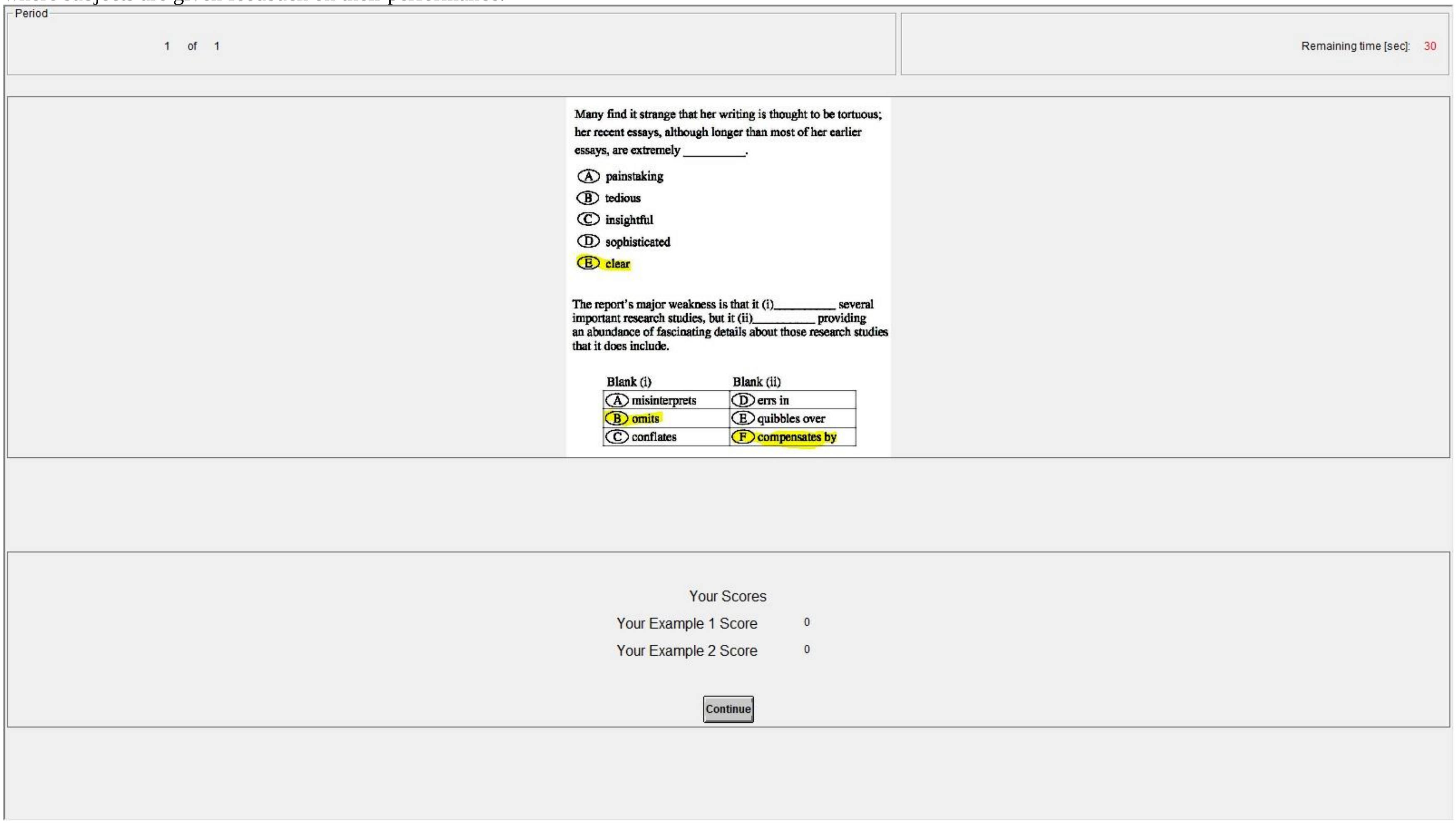
(E) clear

This type of question includes a short text with a blank, indicating that something has been omitted. Select the entry that best completes the text.
Note: You must hit continue below before time elapses to have your answers scored

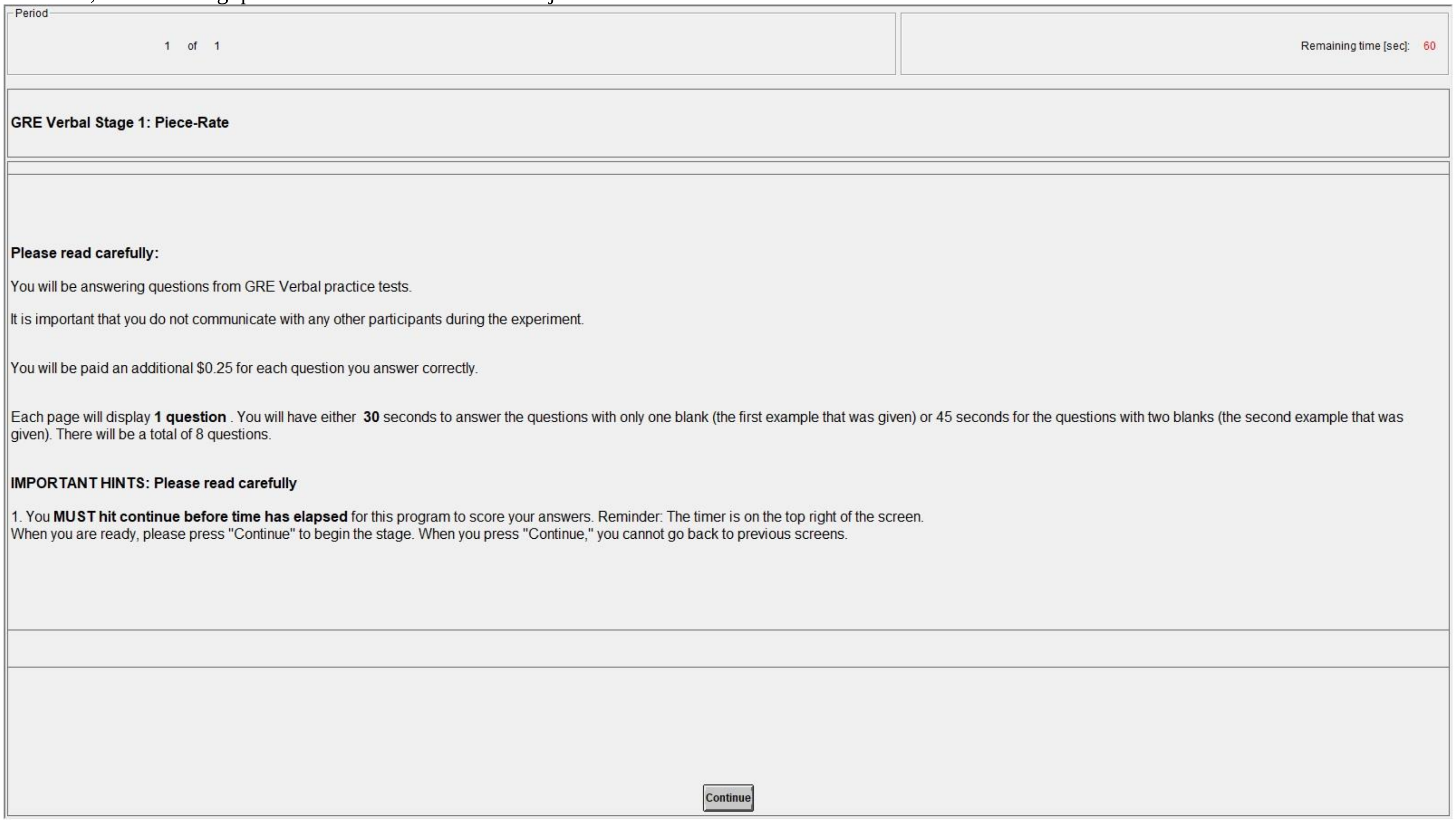
Select Your Answer
- A
- B
- C
- D
- E

Continue

Scores below vary depending on correct answers. Practice round is not formally recorded nor counted towards payoff. The practice rounds are the only time where subjects are given feedback on their performance.



GRE Verbal Round 1 Instructions:
Note: Odd-numbered groups will begin with the piece-rate payoff structure and even-numbered groups will begin with the tournament payoff structure. However, the following questions are the same for all subjects.


Period

1 of 1

Remaining time [sec]: 60

GRE Verbal Stage 1: Piece-Rate

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

It is important that you do not communicate with any other participants during the experiment.

You will be paid an additional $0.25 for each question you answer correctly.

Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Period

1 of 1

Remaining time [sec]: 60

GRE Verbal Stage 1: Tournament

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

It is important that you do not communicate with any other participants during the experiment.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

GRE Verbal Round 1 Questions:

Period

1 of 1

Remaining time [sec]: 30

In the 1950s, the country's inhabitants were _______: most of them knew very little about foreign countries.

(A) partisan

(B) erudite

(C) insular

(D) cosmopolitan

(E) imperturbable

Question 1:

This type of question includes a short text with a blank, indicating that something has been omitted. Select the entry that best completes the text.
Note: You must hit continue below before time elapses to have your answers scored

Select Your Answer
- A
- B
- C
- D
- E

Continue

Period

1 of 1

The playwright's approach is (i)__________ in that her works (ii)__________ the theatrical devices normally used to create drama on the stage.

Blank (i)	Blank (ii)
(A) pedestrian	(D) jettison
(B) startling	(E) experiment with
(C) celebrated	(F) distill

Question 8:

The following question includes a short text with two or three blanks, each blank indicating that something has been omitted. Select one entry for each blank from the corresponding column of choices. Fill all blanks in the way that best completes the text.
Note: You must hit continue below before time elapses to have your answers scored

Answer for (i)
- A
- B
- C

Answer for (ii)
- D
- E
- F

Continue

Period

1 of 1

Remaining time [sec]: 28

GRE Verbal Stage 1 Completion

Congratulations on finishing GRE Verbal Stage 1 of this experiment.

Hit continue when you are ready to move on.

Continue

GRE Verbal Round 2 Instructions:
Note: Odd-numbered groups will continue with the tournament payoff structure and even-numbered groups will continue with the piece-rate payoff structure. However, the following questions are the same for all subjects.

Period

1 of 1

Remaining time [sec]: 60

GRE Verbal Stage 2: Tournament

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Period

1 of 1

GRE Verbal Stage 2: Piece-Rate

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

You will be paid an additional $0.25 for each question you answer correctly.

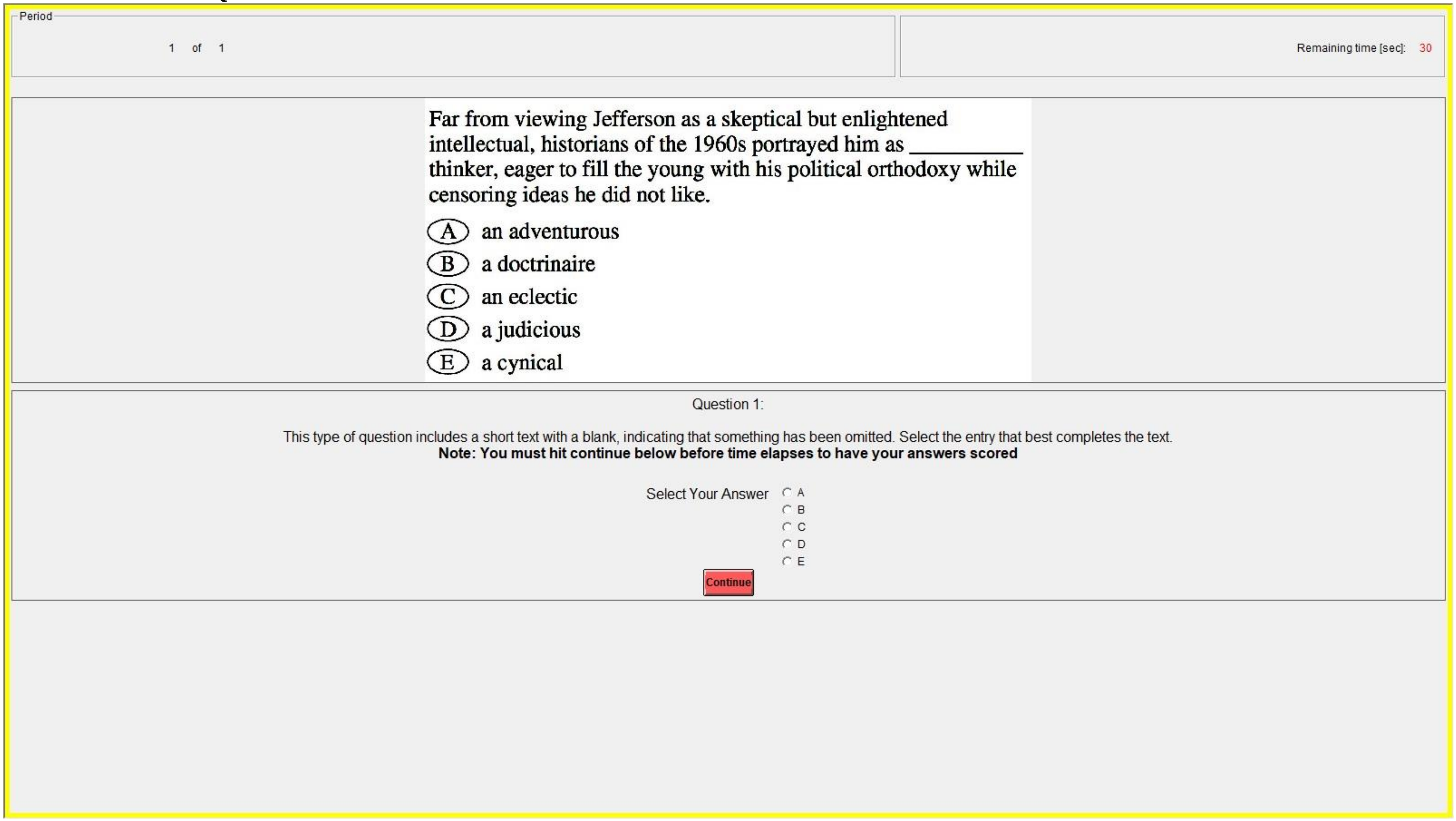
Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

GRE Verbal Round 2 Questions:



Period

1 of 1

Remaining time [sec]: 45

In her later years, Bertha Pappenheim was an apostle of noble but already (i)__________ notions, always respected for her integrity, her energy, and her resolve but increasingly out of step and ultimately (ii)__________ even her own organization.

Blank (i)	Blank (ii)
(A) anachronistic	(D) emulated by
(B) accepted	(E) appreciated by
(C) exotic	(F) alienated from

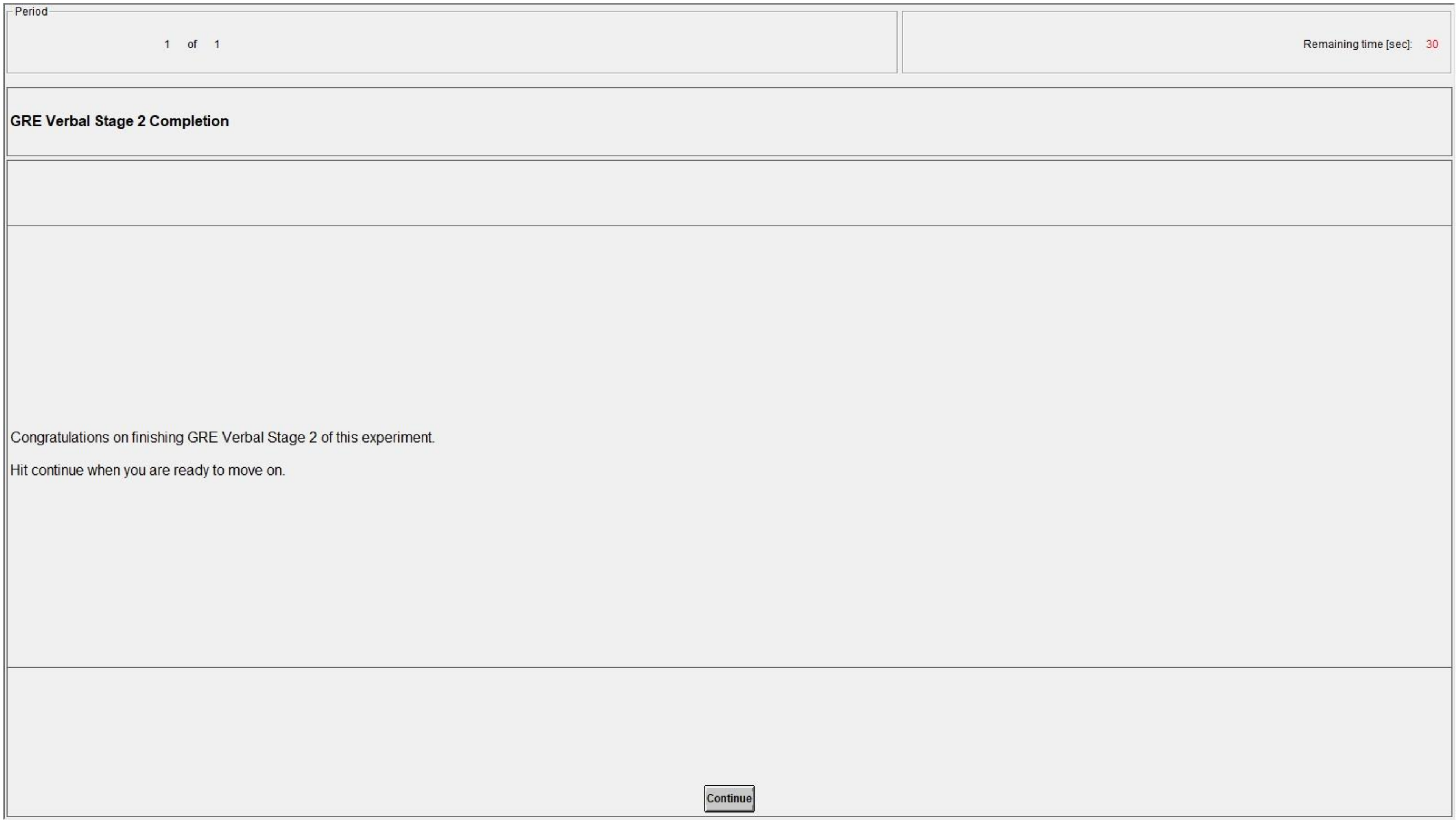
Question 8:

The following question includes a short text with two or three blanks, each blank indicating that something has been omitted. Select one entry for each blank from the corresponding column of choices. Fill all blanks in the way that best completes the text.
Note: You must hit continue below before time elapses to have your answers scored

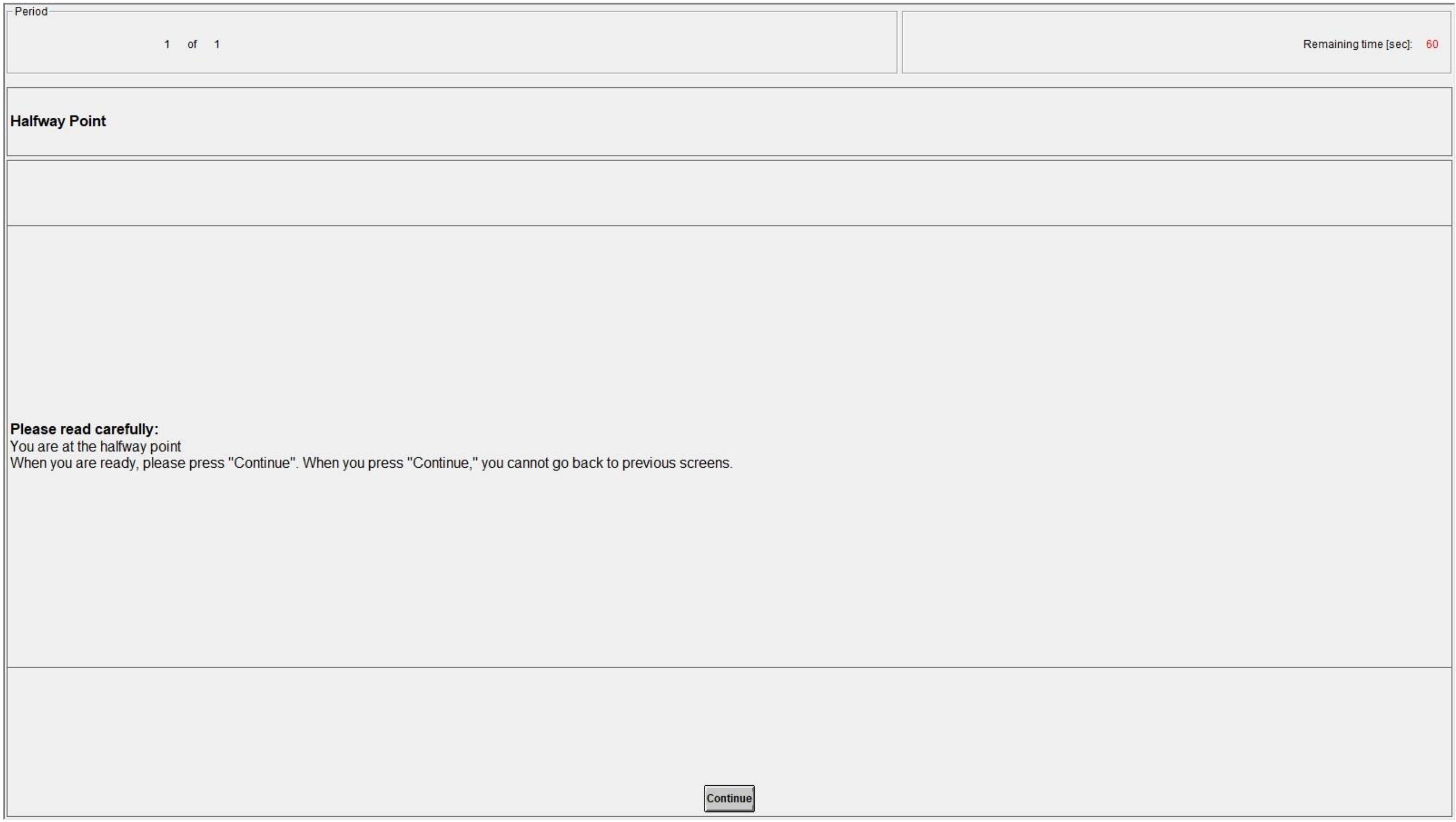
Answer for (i) A B C

Answer for (ii) D E F

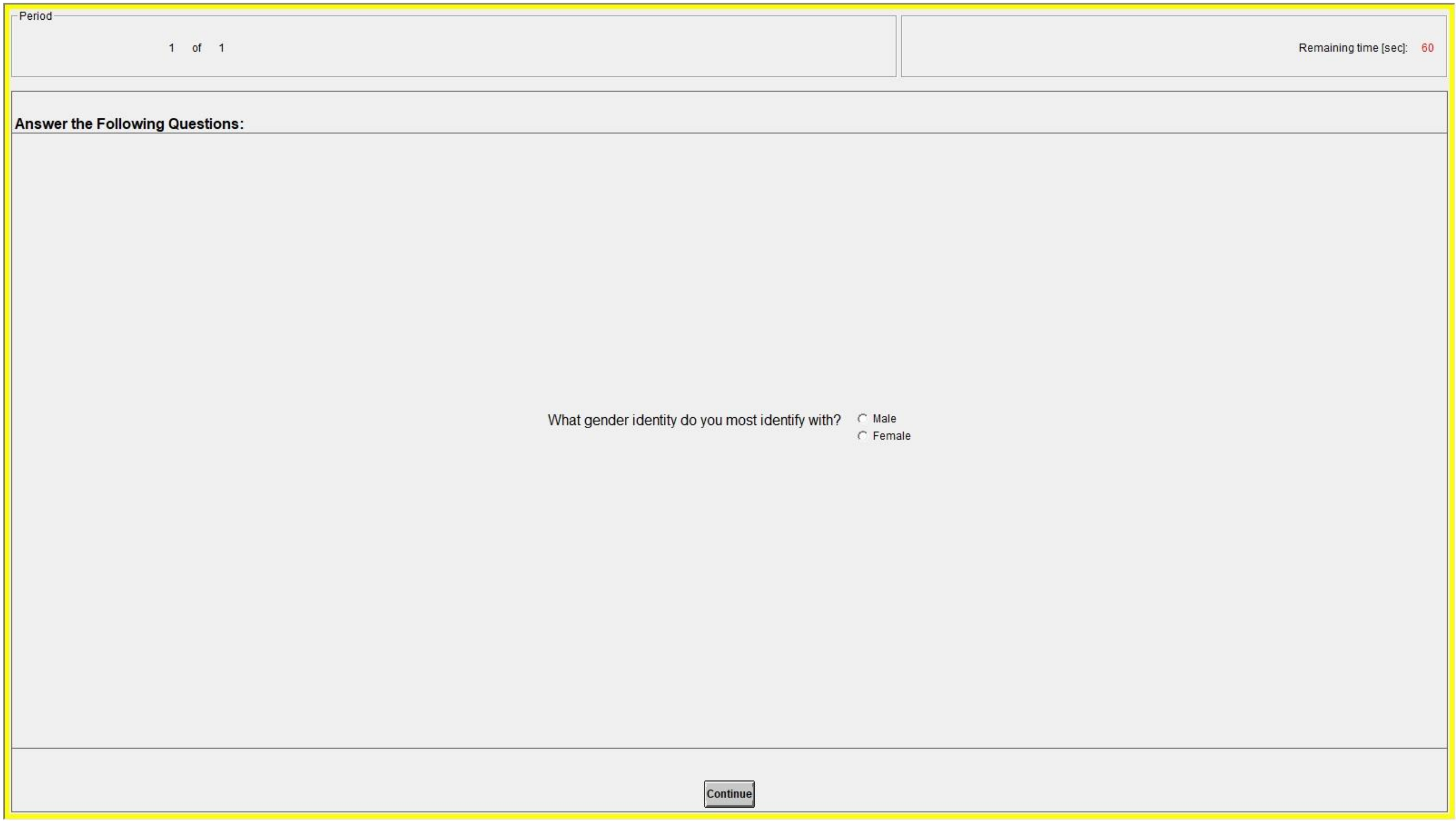
Continue

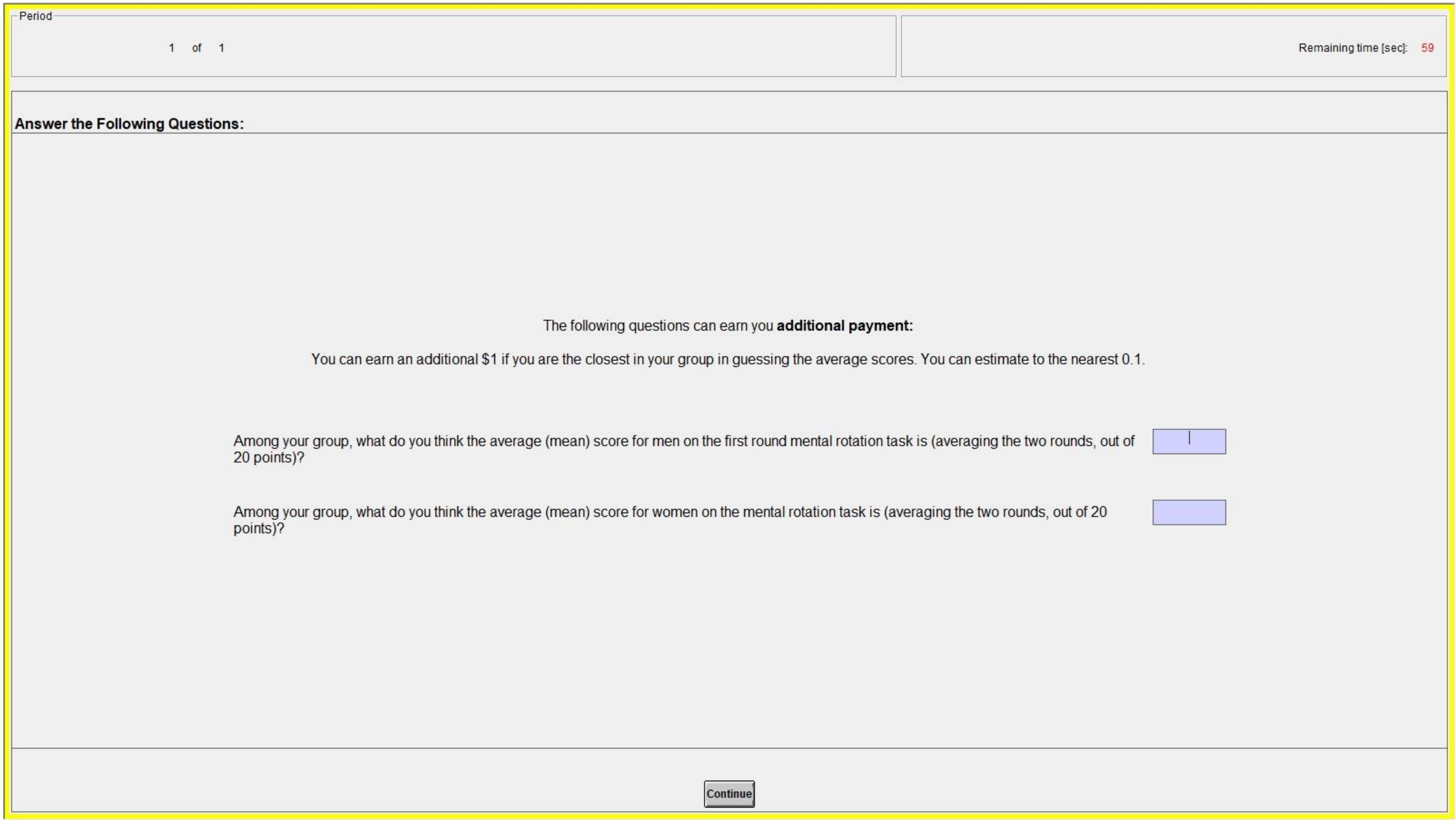


Control Group:
If subjects are in the control group, they see the screen below before proceeding to the next stage

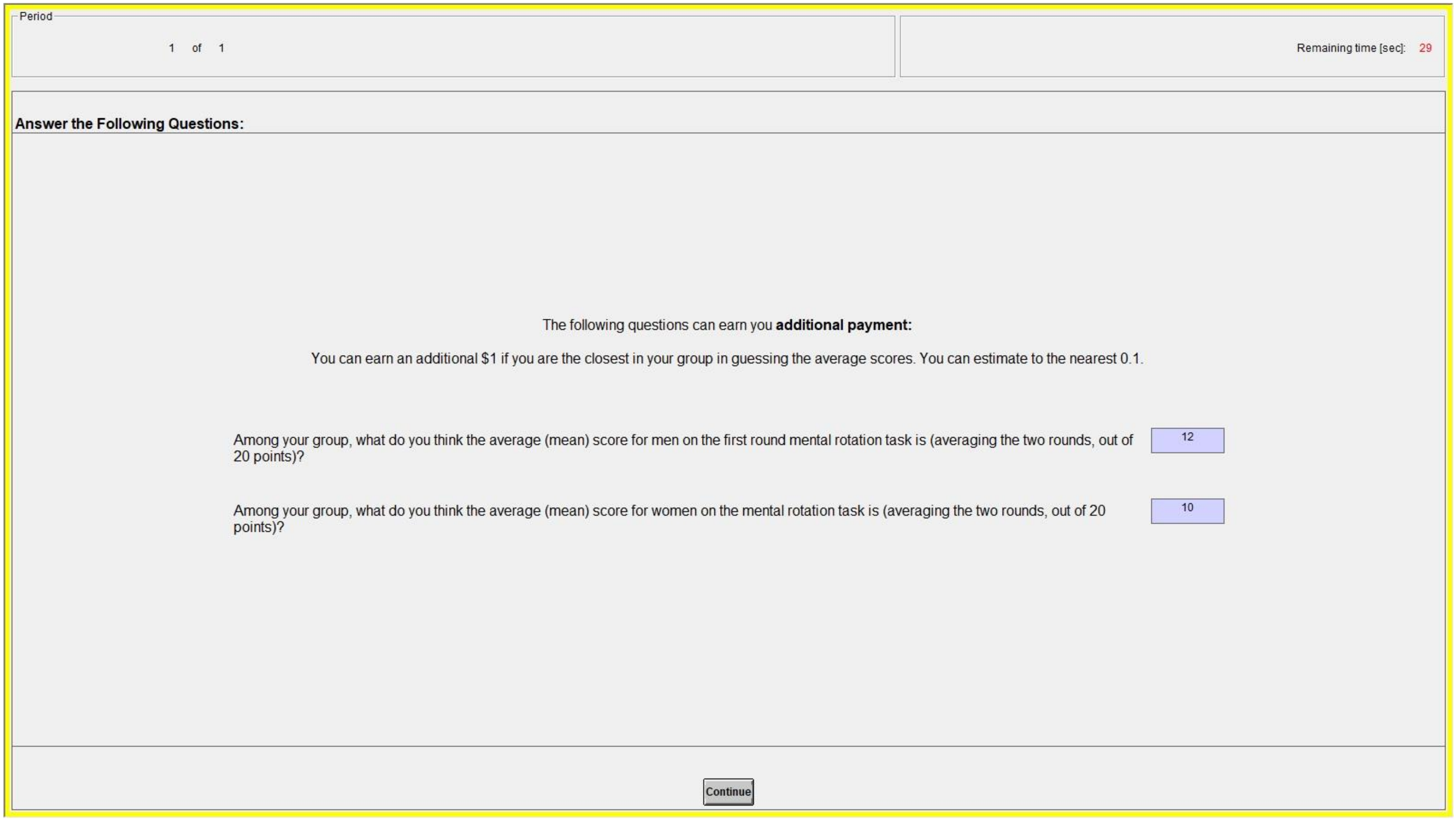


Treatment 1: Perception Treatment
If subjects are in Treatment 1, the perception treatment, they will see the following screens and be asked to answer the following:

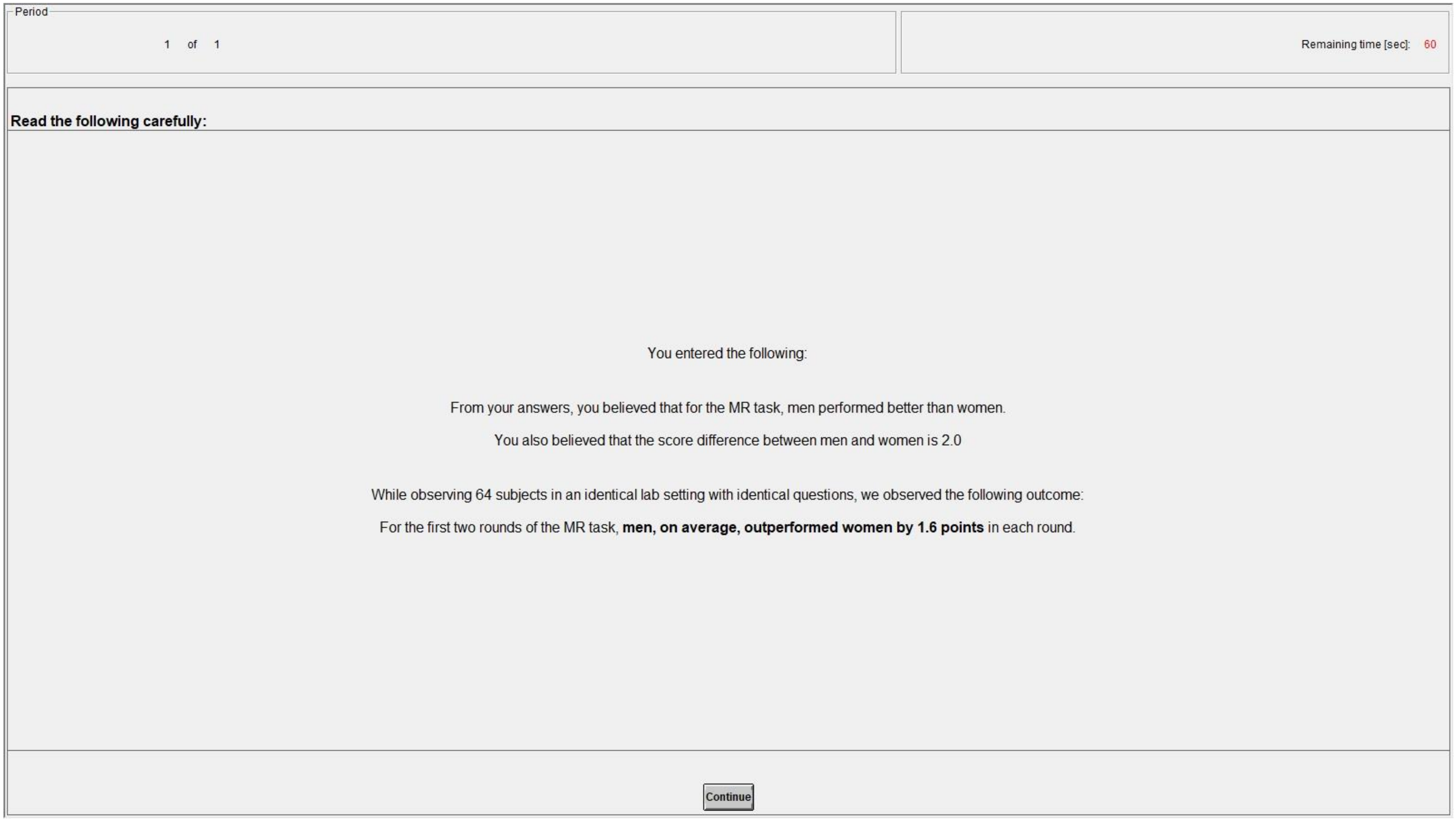

Period
1 of 1
Remaining time [sec]: 60
Answer the Following Questions:
What gender identity do you most identify with?
Male
Female
Continue



Below is an example of a subject who entered in “12” and “10” for the preceding screen:



Treatment 1 and Treatment 2 Interaction: Perception Treatment and Update Treatment
If subjects are exposed to both treatment 1 (perception treatment) and treatment 2 (update treatment), they will see the above screens and the screen below. Below is an example of a subject who entered in “12” and “10” for the preceding screen. The belief in the score difference will vary depending on responses:
If subjects are exposed to treatment 2 (update treatment) exclusively, they will only see the phrase “While observing … in each round.” on this page



Mental Rotation Round 3 Instructions:

Note: Odd-numbered groups will begin with the piece-rate payoff structure and even-numbered groups will begin with the tournament payoff structure. However, the following questions are the same for all subjects.

Period

1 of 1

Remaining time [sec]: 59

Mental Rotation Stage 3: Piece-Rate

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

You will be paid an additional $0.1 for each question you answer correctly.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.

2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.

3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Mental Rotation Round 3 Questions:

Period

1 of 1

Remaining time [sec]: 58

Mental Rotation Stage 3: Tournament

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.

2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.

3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

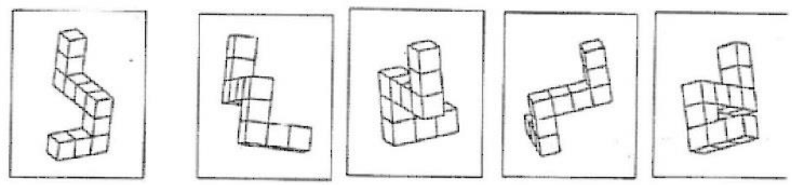
Continue

Period

1 of 1

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 1:



1 **2** **3** **4**

Question 1: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q1 Entry 1:

Q1 Entry 2:

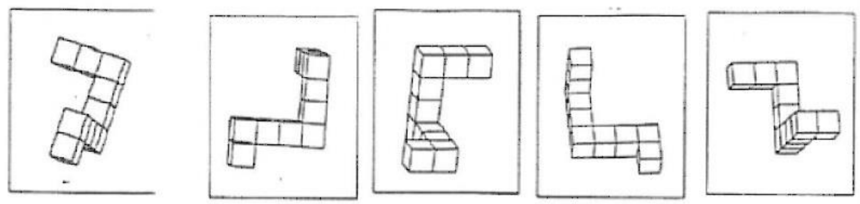
Continue

Period

1 of 1

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 20:



1 2 3 4

Question 20: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q20 Entry 1:

Q20 Entry 2:

Continue

Period

1 of 1

Remaining time [sec]: 25

Mental Rotation Stage 3 Completion

Congratulations on finishing Mental Rotation Stage 3 of this experiment.

Hit continue when you are ready to move on.

Continue

Mental Rotation Round 4 Instructions:

Note: Odd-numbered groups will continue with the tournament payoff structure and even-numbered groups will continue with the piece-rate payoff structure. However, the following questions are the same for all subjects.

Period

1 of 1

Mental Rotation Stage 4: Tournament

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.

2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.

3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Period

1 of 1

Remaining time [sec]: 60

Mental Rotation Stage 4: Piece-Rate

Please read carefully:

You will be conducting mental rotation tasks.

It is important that you do not communicate with any other participants during the experiment.

You will be paid an additional $0.1 for each question you answer correctly.

Each page will display **1 question** . You will have **20 seconds** to answer the question on the page. There will be a total of 20 questions.

IMPORTANT HINTS: Please read carefully

1. This is designed to have time limitations. Make your choices quickly and do not be too concerned if you didn't finish in time. Focus on the next question quickly.

2. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.

3. You can hit the TAB button after entering in Entry 1 to enter in Entry 2 and to go back to Entry 1 if you made a mistake. This saves a mouse click and allows you to enter in your entries quicker.

When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

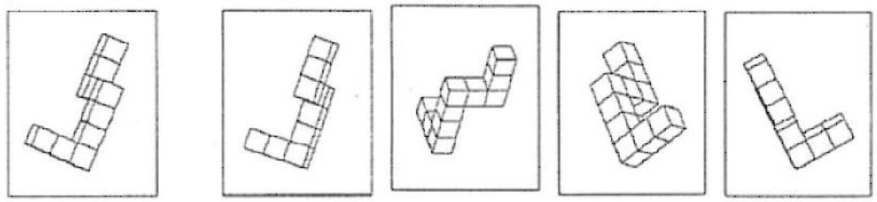
Mental Rotation Round 4 Questions:

Period

1 of 1

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 1:



1 2 3 4

Question 1: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q1 Entry 1:

Q1 Entry 2:

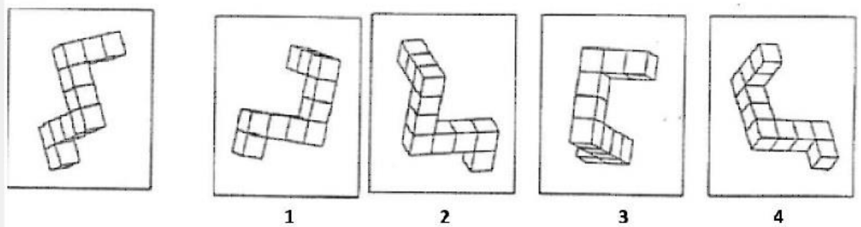
Continue

Period

1 of 1

Please enter two answers for each question. YOU MUST HIT CONTINUE before time has elapsed to have your answers scored:

Question 20:



Question 20: Please Enter in Answers Below. Note: You must hit continue below before time elapses to have your answers scored

Q20 Entry 1:

Q20 Entry 2:

Continue

Period

1 of 1

Remaining time [sec]: 30

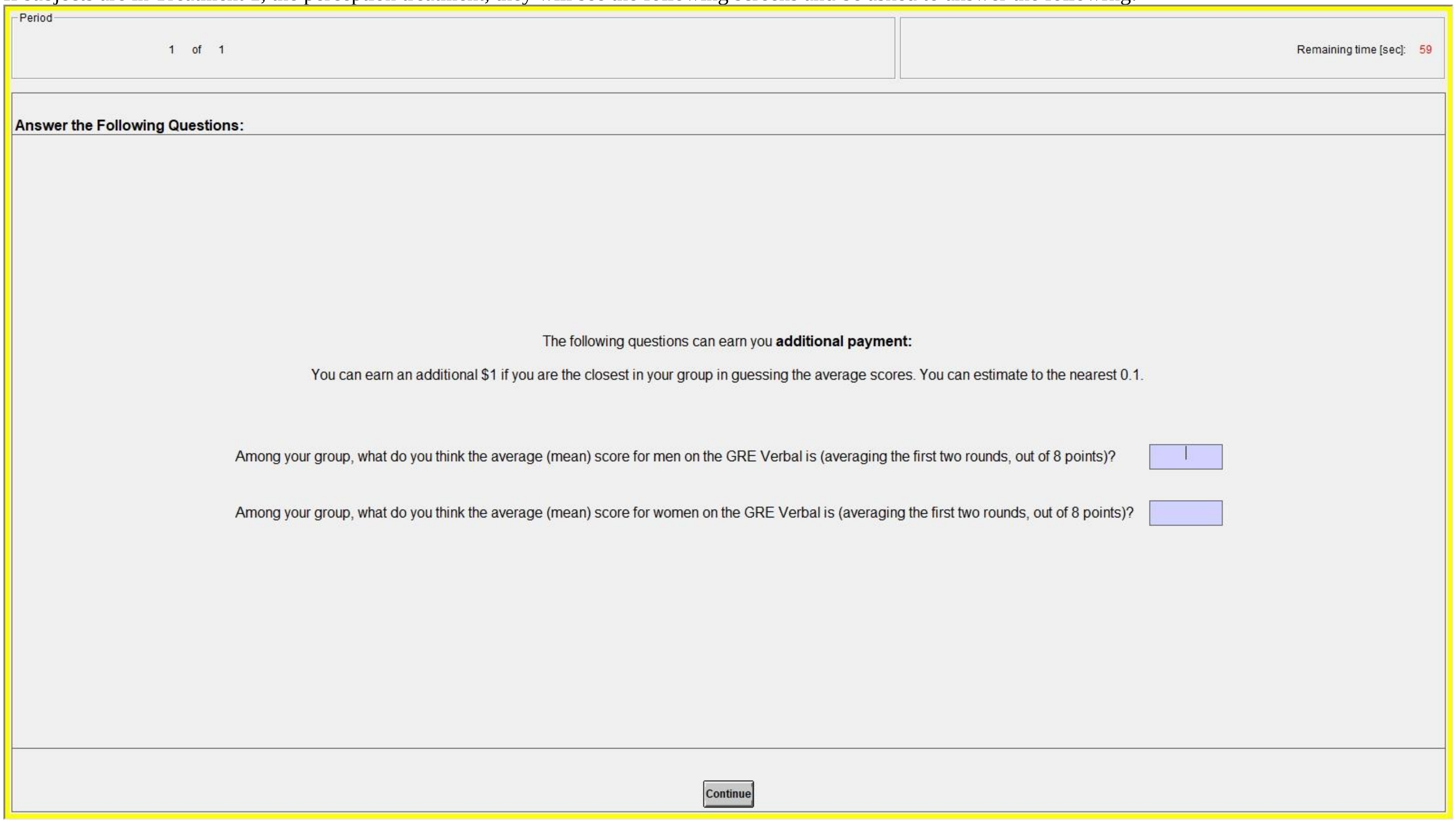
Mental Rotation Stage 4 Completion

Congratulations on finishing Mental Rotation Stage 4 of this experiment.

Hit continue when you are ready to move on.

Continue

Treatment 1: Perception Treatment
If subjects are in Treatment 1, the perception treatment, they will see the following screens and be asked to answer the following:


Period
1 of 1
Remaining time [sec]: 59

Answer the Following Questions:

The following questions can earn you **additional payment:**

You can earn an additional $1 if you are the closest in your group in guessing the average scores. You can estimate to the nearest 0.1.

Among your group, what do you think the average (mean) score for men on the GRE Verbal is (averaging the first two rounds, out of 8 points)?

Among your group, what do you think the average (mean) score for women on the GRE Verbal is (averaging the first two rounds, out of 8 points)?

Continue

Below is an example of a subject who entered in “5” and “7” for the preceding screen:

Period

1 of 1

Remaining time [sec]: 38

Answer the Following Questions:

The following questions can earn you **additional payment:**

You can earn an additional $1 if you are the closest in your group in guessing the average scores. You can estimate to the nearest 0.1.

Among your group, what do you think the average (mean) score for men on the GRE Verbal is (averaging the first two rounds, out of 8 points)? 5

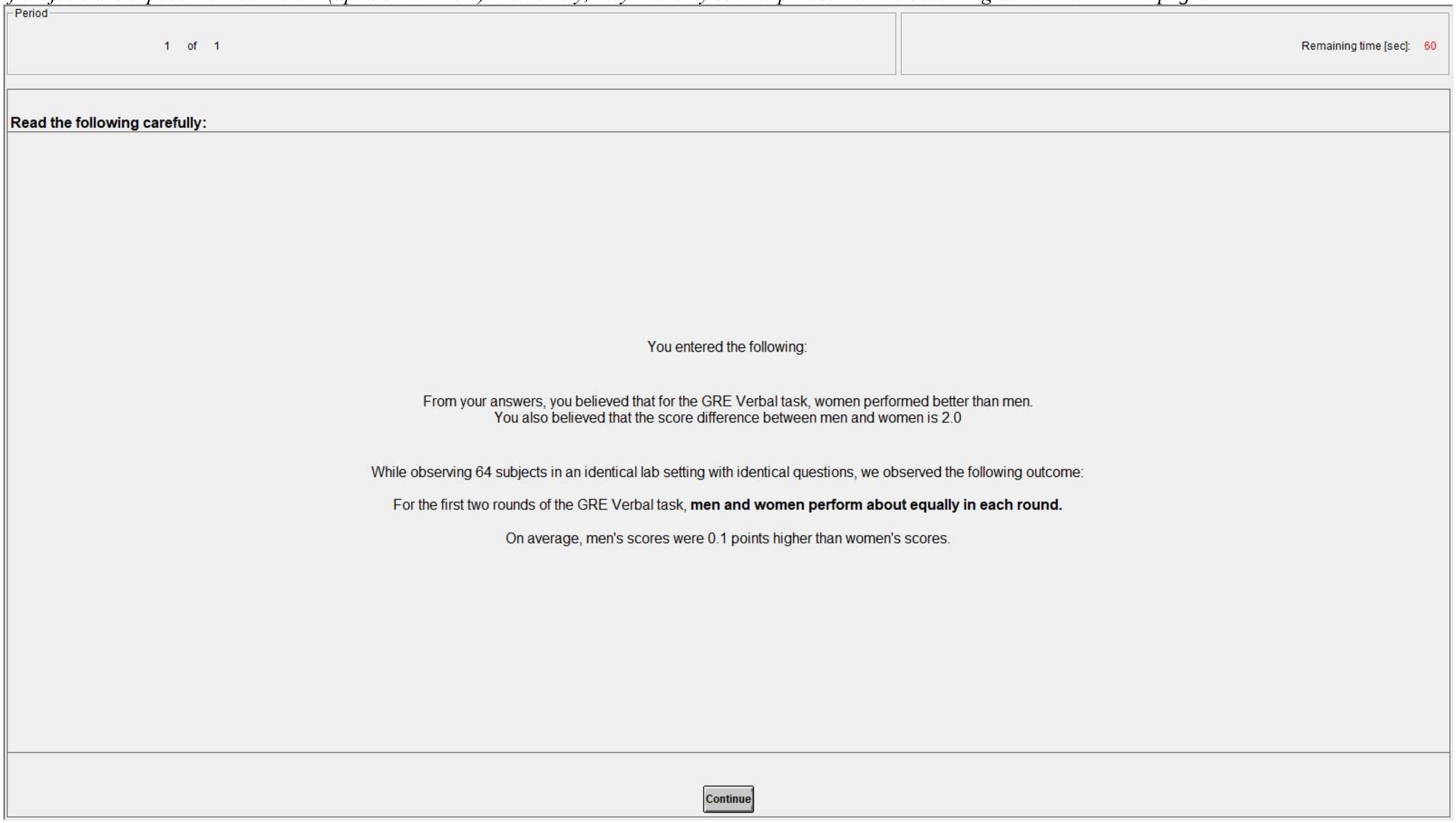
Among your group, what do you think the average (mean) score for women on the GRE Verbal is (averaging the first two rounds, out of 8 points)? 7

Continue

Treatment 1 and Treatment 2 Interaction: Perception Treatment and Update Treatment

If subjects are exposed to both treatment 1 (perception treatment) and treatment 2 (update treatment), they will see the above screens and the screen below. Below is an example of a subject who entered in “5” and “7” for the preceding screen. The belief in the score difference will vary depending on responses:

If subjects are exposed to treatment 2 (update treatment) exclusively, they will only see the phrase “While observing … scores.” in this page.



GRE Verbal Round 3 Instructions:

Note: Odd-numbered groups will begin with the piece-rate payoff structure and even-numbered groups will begin with the tournament payoff structure. However, the following questions are the same for all subjects.

Period

1 of 1

Remaining time [sec]: 60

GRE Verbal Stage 3: Piece-Rate

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

It is important that you do not communicate with any other participants during the experiment.

You will be paid an additional $0.25 for each question you answer correctly.

Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Period

1 of 1

Remaining time [sec]: 60

GRE Verbal Stage 3: Tournament

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

It is important that you do not communicate with any other participants during the experiment.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

GRE Verbal Round 3 Questions:

Period

1 of 1

Remaining time [sec]: 30

The reception given to Kimura's radical theory of molecular evolution shows that when __________ fights orthodoxy to a draw, then novelty has seized a good chunk of space from convention.

(A) imitation

(B) reaction

(C) dogmatism

(D) invention

(E) caution

Question 1:

This type of question includes a short text with a blank, indicating that something has been omitted. Select the entry that best completes the text.
Note: You must hit continue below before time elapses to have your answers scored

Select Your Answer
- A
- B
- C
- D
- E

Continue

Period

1 of 1

A newly published, laudatory biography of George Bernard Shaw fails, like others before it, to capture the essence of his personality: the more he is (i) __________, the more his true self seems to (ii) __________.

Blank (i)	Blank (ii)
(A) discussed	(D) disappear
(B) disparaged	(E) emerge
(C) disregarded	(F) coalesce

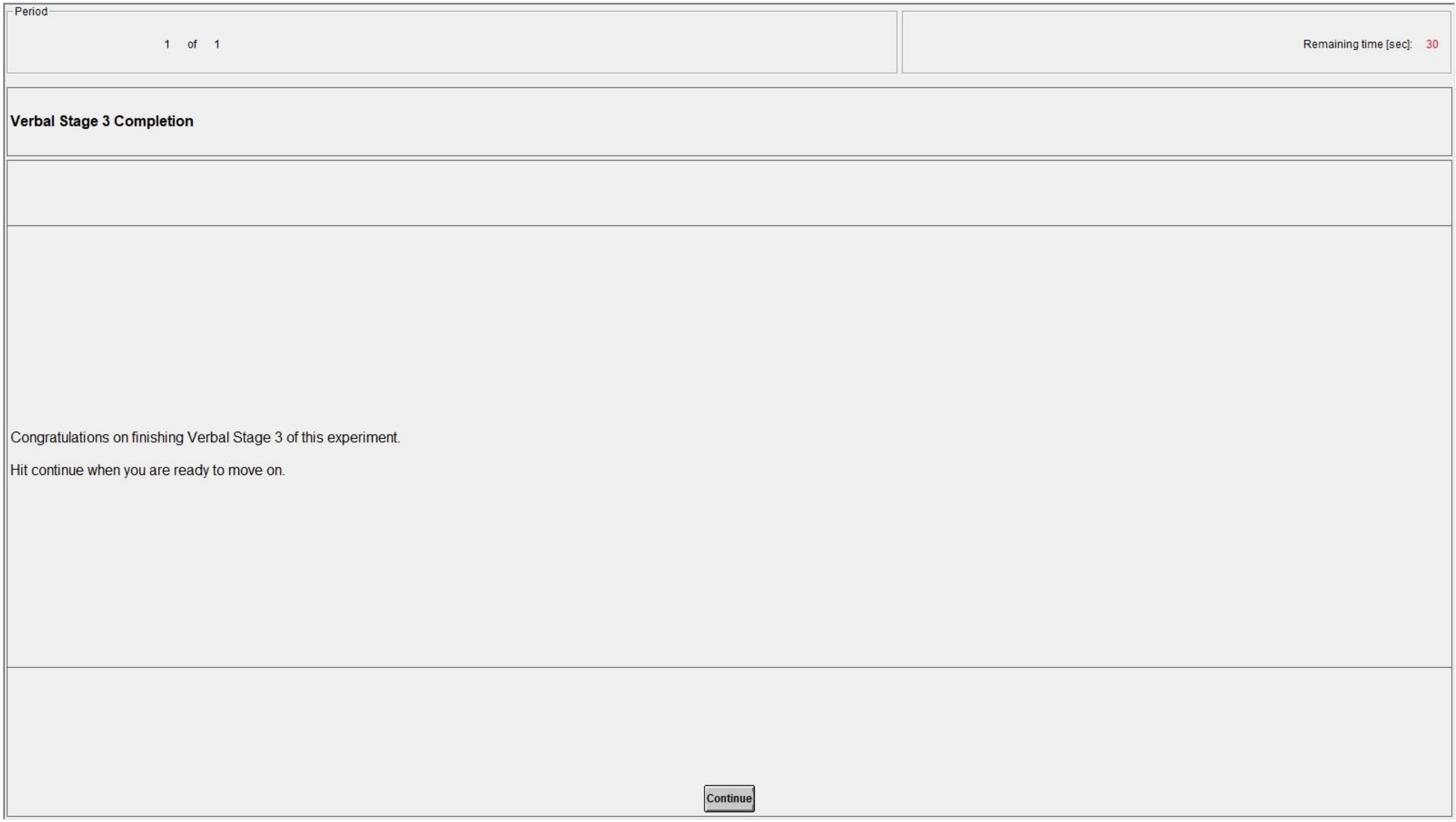
Question 8:

The following question includes a short text with two or three blanks, each blank indicating that something has been omitted. Select one entry for each blank from the corresponding column of choices. Fill all blanks in the way that best completes the text.
Note: You must hit continue below before time elapses to have your answers scored

Answer for (i)
- A
- B
- C

Answer for (ii)
- D
- E
- F

Continue



GRE Verbal Round 4 Instructions:

Note: Odd-numbered groups will continue with the tournament payoff structure and even-numbered groups will continue with the piece-rate payoff structure. However, the following questions are the same for all subjects.

Period

1 of 1

Remaining time [sec]: 60

GRE Verbal Stage 4: Tournament

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

It is important that you do not communicate with any other participants during the experiment.

The style of payment here is **tournament style** . If you score the highest in your group, you will earn an additional $4.

Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

Period

1 of 1

GRE Verbal Stage 4: Piece-Rate

Please read carefully:

You will be answering questions from GRE Verbal practice tests.

It is important that you do not communicate with any other participants during the experiment.

You will be paid an additional $0.25 for each question you answer correctly.

Each page will display **1 question** . You will have either **30** seconds to answer the questions with only one blank (the first example that was given) or 45 seconds for the questions with two blanks (the second example that was given). There will be a total of 8 questions.

IMPORTANT HINTS: Please read carefully

1. You **MUST hit continue before time has elapsed** for this program to score your answers. Reminder: The timer is on the top right of the screen.
When you are ready, please press "Continue" to begin the stage. When you press "Continue," you cannot go back to previous screens.

Continue

GRE Verbal Round 4 Questions:

Period

1 of 1

Remaining time [sec]: 30

Mechanisms develop whereby every successful species can __________ its innate capacity for population growth with the constraints that arise through its interactions with the natural environment.

(A) enhance
(B) replace
(C) produce
(D) surpass
(E) reconcile

Question 1:

This type of question includes a short text with a blank, indicating that something has been omitted. Select the entry that best completes the text.
Note: You must hit continue below before time elapses to have your answers scored

Select Your Answer
- A
- B
- C
- D
- E

Continue

Period

The (i)__________ nature of classical tragedy in Athens belies the modern image of tragedy: in the modern view tragedy is austere and stripped down, its representations of ideological and emotional conflicts so superbly compressed that there's nothing (ii)__________ for time to erode.

Blank (i)	Blank (ii)
(A) unadorned	(D) inalienable
(B) harmonious	(E) exigent
(C) multifaceted	(F) extraneous

Question 8:

The following question includes a short text with two or three blanks, each blank indicating that something has been omitted. Select one entry for each blank from the corresponding column of choices. Fill all blanks in the way that best completes the text.

Note: You must hit continue below before time elapses to have your answers scored

Answer for (i) ○ A ○ B ○ C

Answer for (ii) ○ D ○ E ○ F

Continue

Period

1 of 1

Remaining time [sec]: 30

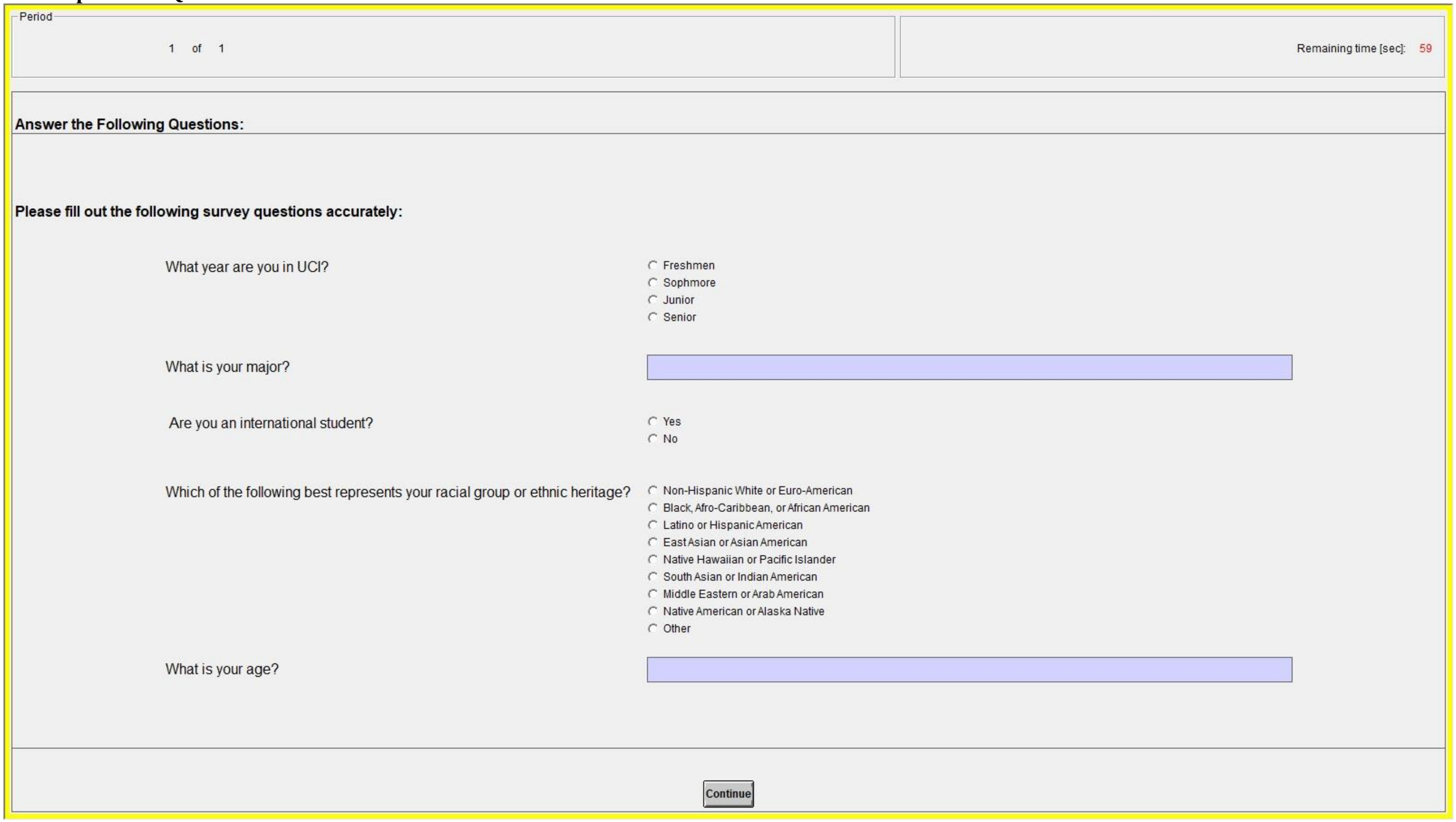
Verbal Stage 4 (Final Stage) Completion

Congratulations on finishing the experiment.

Hit continue to fill out a brief questionaire.

Continue

Post-Experiment Questionnaire:



Period

1 of 1

Remaining time [sec]: 59

Answer the Following Questions:

Please fill out the following survey questions accurately:

What year are you in UCI?

- Freshmen
- Sophmore
- Junior
- Senior

What is your major?

Are you an international student?

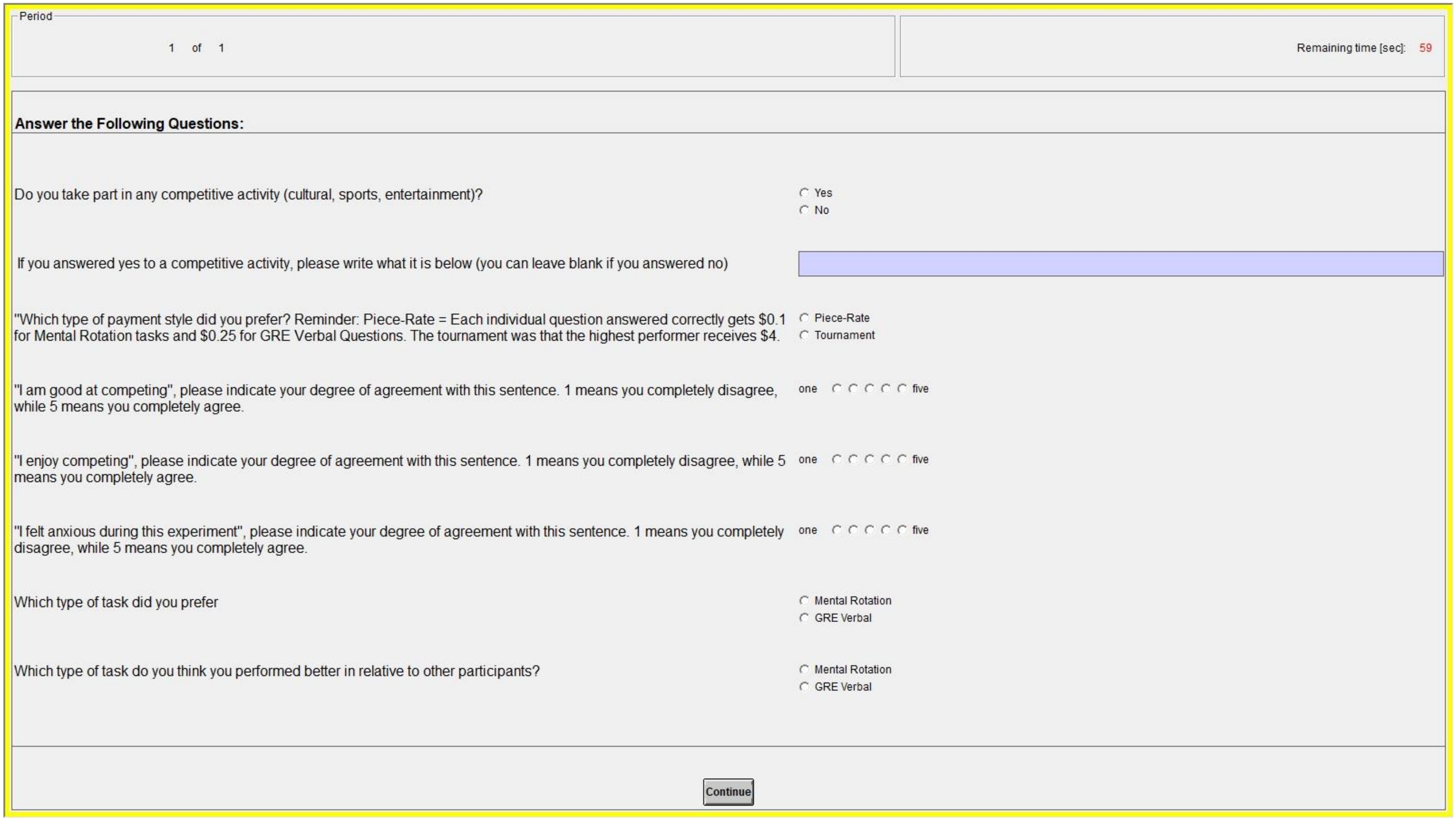
- Yes
- No

Which of the following best represents your racial group or ethnic heritage?

- Non-Hispanic White or Euro-American
- Black, Afro-Caribbean, or African American
- Latino or Hispanic American
- East Asian or Asian American
- Native Hawaiian or Pacific Islander
- South Asian or Indian American
- Middle Eastern or Arab American
- Native American or Alaska Native
- Other

What is your age?

Continue



Period

1 of 1

Remaining time [sec]: 59

Answer the Following Questions:

Do you take part in any competitive activity (cultural, sports, entertainment)?

- Yes
- No

If you answered yes to a competitive activity, please write what it is below (you can leave blank if you answered no)

"Which type of payment style did you prefer? Reminder: Piece-Rate = Each individual question answered correctly gets $0.1 for Mental Rotation tasks and $0.25 for GRE Verbal Questions. The tournament was that the highest performer receives $4.

- Piece-Rate
- Tournament

"I am good at competing", please indicate your degree of agreement with this sentence. 1 means you completely disagree, while 5 means you completely agree.

one ○ ○ ○ ○ ○ five

"I enjoy competing", please indicate your degree of agreement with this sentence. 1 means you completely disagree, while 5 means you completely agree.

one ○ ○ ○ ○ ○ five

"I felt anxious during this experiment", please indicate your degree of agreement with this sentence. 1 means you completely disagree, while 5 means you completely agree.

one ○ ○ ○ ○ ○ five

Which type of task did you prefer

- Mental Rotation
- GRE Verbal

Which type of task do you think you performed better in relative to other participants?

- Mental Rotation
- GRE Verbal

Continue

Period

1 of 1

Remaining time [sec]: 46

Answer the Following Questions:

The following questions can earn you **additional payment:**

Questions 1-2 and Questions 5-6 can earn you an additional $0.25 each if you are the closest in your group (for each set)

Among your group, what do you think the average (mean) score for men on the mental rotation task is (averaging all four rounds, out of 20 points)?

Among your group, what do you think the average (mean) score for women on the mental rotation task is (averaging all four rounds, out of 20 points)?

Among all participants today, what do you think the average (mean) score for men on the mental rotation task is (averaging all four rounds, out of 20 points)?

Among all participants today, what do you think the average (mean) score for women on the mental rotation task is (averaging all four rounds, out of 20 points)?

Among your group, what do you think the average (mean) score for men on the GRE Verbal is (averaging all four rounds, out of 8 points)?

Among your group, what do you think the average (mean) score for women on the GRE Verbal is (averaging all four rounds, out of 8 points)?

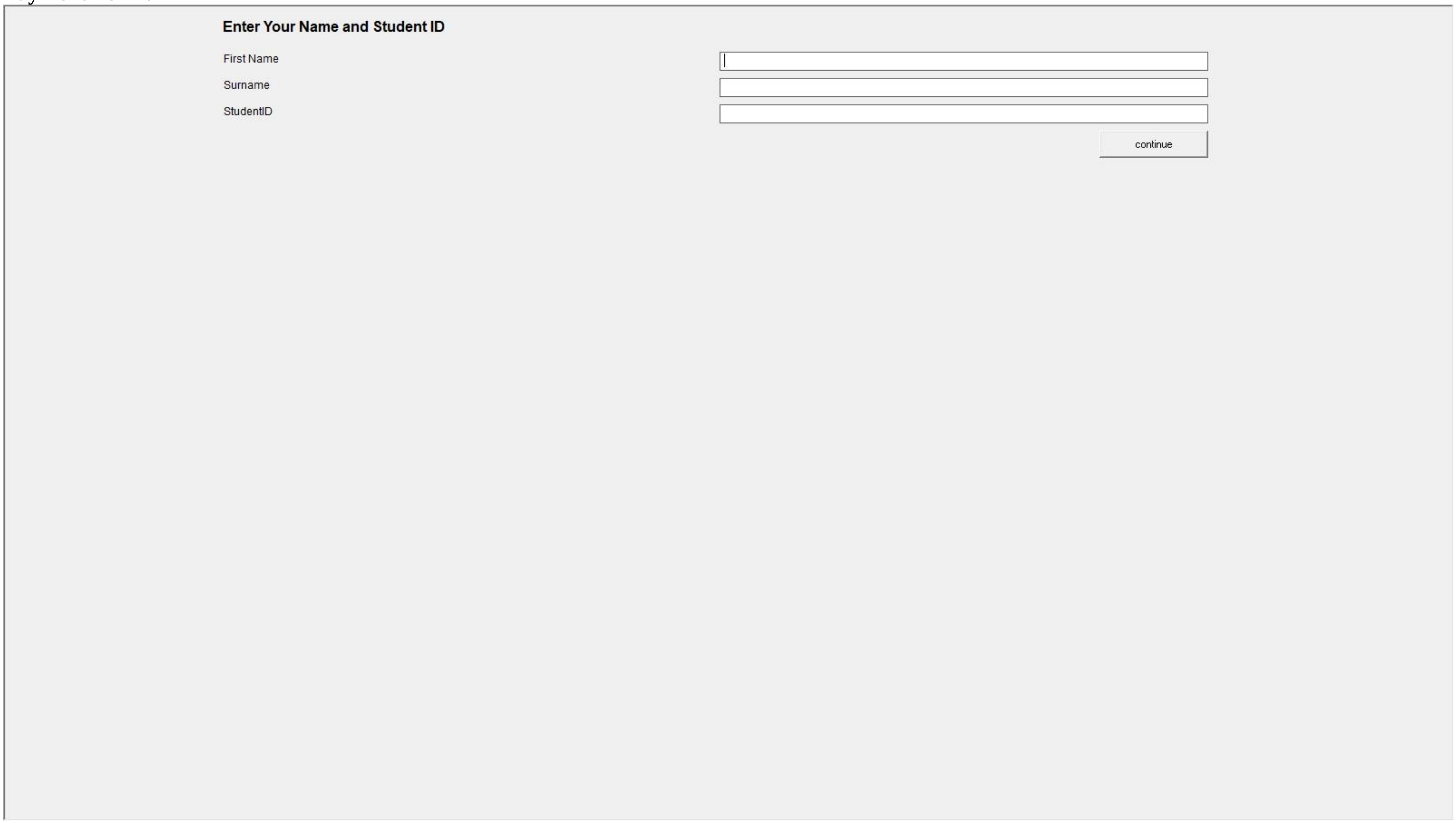
Among all participants today, what do you think the average (mean) score for men on the GRE Verbal is (averaging all four rounds, out of 8 points)?

Among all participants today, what do you think the average (mean) score for women on the GRE Verbal is (averaging all four rounds, out of 8 points)?

Continue

Experiment Completion
The experiment has concluded. You will be asked to fill out a form that will generate a payment file and let you know your total payment from this session.

Payment Form:



Enter Your Name and Student ID

First Name

Surname

StudentID

continue

Show-Up Payment	7.00
Variable Amount Earned	18.75
Your Total Earnings	25.75

Appendix E: LODES Data on the Overall "Connectedness" of the Route

The origin-destination pairs in the LODES data can be used to evaluate the effect of the overall connectedness of the Metro system in residency to workplace pairs. To do this, I redefine the panel variable from tracts to all origin-destination tract-pairs of every worker that either lives or works in L.A. County. Then, I can evaluate the effect of a connected route, defined as when the tract of residency and the tract of work both contain a Metro station, on the number of connected jobs. I use the entirely of the LODES data from 2002-2016 to evaluate multiple expansion routes. A proposed specification for this is as follows:

$$Y_{zt} = \alpha + \beta_1 * MetroWork_{zt} * MetroLive_{zt} + \beta_2 * MetroLive_{zt} + \beta_3 * MetroWork_{zt} + \Upsilon_c + \lambda_t + \varepsilon_{ct} \qquad \text{(E.1)}$$

$MetroLive_{zt}$ indexes whether the origin, or place of residency, of an origin-destination pair z contains a Metro station within walking distance in time t and $MetroWork_{zt}$ indexes whether the destination, or place of work, of the origin-destination pair z contains a Metro station within walking distance in time t. When both $MetroLive_{zt}$ and $MetroWork_{zt}$ are equal to one, the origin-destination pair contains a Metro station in both the place of residency and work. Therefore, β1 identifies the overall effect of having a connected route from residency to work on the number and type of jobs. ϒz are dummy variables for all origin-destination pairs and λt are year dummy variables from 2002-2016.

There are 139,588,875 origin-destination commuting pairs rather than by tract to connect both commuting characteristics and the number and type of job (Appendix Table E1). This table shows that the connectedness of the route matters. When origin-destination pairs contains a Metro station in the place of residence and work, on average, the total number of commutes along this route increases by 0.023. However, when the route is not connected but contains either a place of residency with a Metro station or a place of work with a Metro station, there are generally statistically significant decreases in the number of commutes. Thus, these results show that the connectedness of the route matters.

Appendix Table E1: Metro on Number and Type of Jobs (for all Origin-Destinations)

Variables	Total Jobs	Low Inc. Jobs	Med Inc. Jobs	High Inc. Jobs
LiveAtMetro*WorkAtMetro	0.023***	-0.015***	-0.029***	0.067***
	[0.007]	[0.003]	[0.003]	[0.004]
WorkAtMetro	-0.016**	0.008***	-0.014***	-0.011***
	[0.002]	[0.001]	[0.001]	[0.001]
LiveAtMetro	-0.006***	-0.003***	-0.004***	0.001
	[0.001]	[0.001]	[0.001]	[0.001]
constant	0.522***	0.144***	0.190***	0.188***
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	139,588,875	139,588,875	139,588,875	139,588,875
R-Squared	0.823	0.666	0.697	0.762

The unit of observation is all origin-destination pairs with at least one origin or one destination in Los Angeles County. Robust standard errors clustered by origin-destination pairs. *, **, and *** denote 0.1, 0.05, and 0.01 significance levels.