Stigma and Take-Up of Labor Market Assistance: Evidence from Three Experiments*

Adam Osman[†] Jamin D. Speer[‡]

Current Draft: April 2021
First Draft: July 2020

Abstract

Aversion to "stigma" may affect utilization of social programs, but empirical evidence of its importance is scarce. Using three randomized field experiments, we show that stigma can affect consequential labor market decisions. Treatments designed to alleviate stigma concerns had small average effects on take-up. However, we document large and important heterogeneity. Stigma significantly effects the composition of who takes up a program: applicants from treatment differ along several dimensions. On the other hand, "welfare stigma" does not reduce take-up for any group. We conclude that stigmas are stronger for program outcomes (i.e., entry-level jobs) than for entering the programs themselves.

*The authors thank Sarah Wahby, Abdelrahman Nagy, Mohamed Yousef, Mohsen Nagy, Mohamed Omar, Norhan Muhab, Eslam Serrag, Mahmoud Shehata, and Noha Fadl from J-PAL; Eduardo Medina for research assistance; Ismail Habrouk, Nora Abou El Seoud, Farah Osman, Theo Wilson, Julie DeSoto and Hana El Fekky from EFE; Iris Boutros, Arjan De Haan, Rania Fazah and Flaubert Mbiekop from IDRC; Mohamed El Komi and Ebaa El Sayed from AUC; Mahmoud El Bably from DevGate; Hesham El Badawy from JobMaster; and the participants in the study. We thank Kareem Haggag, Ben Harrell, Joaquin Lopez, and anonymous referees for valuable comments. UIUC IRB Approval #16766, Memphis IRB Approval #4166, AEA Trial Registry AEARCTR-0006131. We are thankful for financial support from the International Development Research Centre. Authors retained full intellectual freedom to design, analyze and interpret results. All errors are our own.

†University of Illinois Urbana-Champaign. Email: aosman@illinois.edu

‡University of Memphis. Email: jspeer@memphis.edu

1 Introduction

Take-up of social programs is often low despite large expected benefits (Currie, 2006) and is known to vary across demographic groups (Heckman and Smith, 2004). This is true in both high- and low-income countries (Rinehart and McGuire, 2017). There are many potential explanations, including lack of information, costs of applying for benefits, and stigma (Bhargava and Manoli, 2015; Friedrichsen et al., 2018). Stigma – a disutility associated with participating in certain activities or programs – is widely believed to be a major factor in low take-up of social programs (Moffitt, 1983), but there is limited empirical evidence of its effects.[1]

Job assistance programs suffer from similar take-up issues, but also need to contend with the uncertain labor market returns associated with participating in these types of programs. There is uncertainty about the average treatment effect of programs – with some finding positive effects (Attanasio et al., 2017, 2011; Bandiera et al., 2017) and others no effect (Groh et al., 2016; Hirshleifer et al., 2016) – as well as heterogeneity in who benefits (Acevedo et al., 2020; Card et al., 2018; Kluve et al., 2019; McKenzie, 2017). Given this heterogeneity, it is essential to understand how people select into these programs and how recruitment practices affect the size and makeup of the applicant pool (World Bank Group, 2018).

This paper uses three randomized experiments in Cairo, Egypt, to study the effects of information provision – specifically, information about stigma – on take-up of labor market assistance programs. The first two experiments recruit unemployed youth to a job training program via Facebook and street-level marketing, respectively. The third recruits individuals to attend a job fair using door-to-door outreach. These programs focused on trying to help unemployed and underemployed youth get a job in the formal sector, usually in entry-level positions.

In our three experiments, we randomly vary the "pitch" delivered to individual job-seekers about these free or highly subsidized job assistance programs. The control groups receive basic information about the programs and their potential outcomes,

[1] The term "stigma" is used broadly in the literature. It includes "social stigma"- disutility due to what other people think of one's participation (Major et al., 1998); "personal stigma"- disutility due to how one feels about oneself (Manchester and Mumford, 2010); "ability stigma"- being seen as less able; and "free-rider stigma"- being seen as willing to live off others (Friedrichsen et al., 2018).

while treatment groups receive this plus additional information designed to help individuals overcome worries about any perceived stigmas associated with taking entry-level jobs in low-skill professions.

We find some evidence that stigma can depress take-up on average. This negative average effect is only present in one of the three experiments we run. In the other two experiments, we find average effects that are mostly negative but not statistically different from zero. However, our key finding is that a lack of average impacts of stigma can mask large and important heterogeneity in responses. We find strong evidence that the impacts of stigma surrounding entry-level jobs are heterogeneous, which has important implications for research and for programs' recruitment and effectiveness.

We assess this heterogeneity in two ways. One is through compositional analysis, in which we ask if a treatment aimed at alleviating stigma changes the composition of who participates in a program. Even if such a treatment has no average effect on take-up, it can alter who participates if the effects are heterogeneous. To provide a simple example, if stigma is relevant for men but not for women, then a treatment aimed at overcoming stigma will produce applicants of a different gender mix than the control group will. If we see such a difference in composition, this is evidence that our interventions have heterogeneous impacts by individual characteristics.

This is indeed what we find. In both experiments where we are able to look for heterogeneity, despite negligible average effects, the stigma treatments deliver applicants that are richer, older, and more likely to be currently working, relative to applicants from the control group. In many cases the differences are large and significant.

However, not all heterogeneous impacts may present themselves so clearly. Interventions may affect different groups in different ways that are hard to predict ex-ante, and the potential differences could be multi-dimensional. Thus, we assess heterogeneity in a another way, utilizing machine learning techniques based on Chernozhukov et al. (2020). These methods use generic machine learning algorithms that predict the individual treatment effect for study participants using baseline data on others in the sample. This ensures that the estimated individual treatment effects are not merely a form of data-mining, but are in fact "honest", and are picking up heterogeneity that is partially generalizable. This is because the estimate is based on a training sample and does not include the effect coming from the individual themselves.

Using these methods, we again find strong evidence of heterogeneous effects of stigma. For the same recruitment pitch, the treatment effects for some groups of individuals are significantly different from the treatment effects for others. Some people are responding strongly negatively to our message, while others are responding positively or not at all. We also generate a Lasso-based multi-dimensional index of baseline characteristics that predicts the individual treatment treatment effect. We find that those who apply for the program in the treatment group are more likely to score higher on this index, relative to those who apply for the program in control.

Both methods show strong evidence that stigma surrounding entry-level jobs has heterogeneous effects on take-up. This means that studies finding small or no average impacts of a stigma treatment are not necessarily evidence that stigma is unimportant. Many studies have failed to find much evidence of stigma depressing take-up (see Currie (2006) for a survey), but stigma could still be acting on different groups in ways that cancel each other out.

Stigma could also reduce take-up of these programs for another reason. In addition to stigmas surrounding the outcomes of the programs, there could be a negative stigma associated with taking part in any kind of assistance program itself, particularly one intended for the poor. This "welfare stigma" is the focus of most of the stigma literature in economics (e.g., Moffitt, 1983).

Our last key finding is that stigmas surrounding program *outcomes* seem to be more important than stigma surrounding program participation itself. We find little evidence for the importance of welfare stigma in this context. Telling recruits that the program is subsidized "to help those in financial hardship" has no significant effect on application rates or program composition, and we do not find significant evidence of heterogeneity in these effects.

We make several important contributions to the literature. First, we show that social image concerns surrounding entry-level jobs can affect labor market decisions and investments. This is important because these jobs are typically the stepping stone to future employment (Groes et al., 2015; Sicherman and Galor, 1990). The literature on social image (surveyed in Bursztyn and Jensen (2017)) studies how people's self-perception and society's perception of them can influence their behavior, both for good and for ill. People will pay to project a certain image to others (Bursztyn et al., 2016; Cruces et al., 2013), and willingness to do so is heterogeneous (Friedrichsen and Engelmann, 2018). Concerns about social image can impact educational investments (Bursztyn and Jensen, 2015), attitudes about gender and work (Bursztyn et al., 2020), and personal financial decisions (Ghosal et al., 2020). Our study draws a direct connection between social image concerns and decisions about jobs and unemployment.

We also show that there is substantial heterogeneity in how stigma affects people's labor market decisions. Maturity, work experience, and family income seem to mitigate the negative effects of stigma in our experiments. It is important for governments, program administrators, and researchers to be aware of potential heterogeneous effects of various recruitment methods. Different recruiting tools can alter the composition of the participant pool even if they have no effect on overall take-up, which could alter the effectiveness of programs in important ways (Card et al., 2018). Depending on the program's goals, this may also affect how well-targeted the assistance is.

Finally, we show clear evidence that welfare stigma is not important in this context. While welfare stigma is often assumed to exist (Besley and Coate, 1995; Moffitt, 1983; Stuber and Schlesinger, 2006), it is difficult to distinguish from transaction costs (Currie, 2006), and there is little empirical evidence of its importance (Currie, 2003; Remler and Glied, 2003; Schofield et al., 2019; Stuber and Schlesinger, 2006). A recent lab experiment by Friedrichsen et al. (2018) finds that stigma reduces take-up of a welfare-like benefit. On the other hand, the interventions by Bhargava and Manoli (2015) aimed at overcoming welfare stigma are ineffective. Our setting is one in which the potential for real-world stigma is more salient, but our results provide clean evidence that welfare stigma nonetheless has minimal effects on program take-up and composition. As far as we are aware, ours is the first experimental evidence on welfare stigma in a developing country setting.

The paper proceeds as follows. Section 2 discusses the local context of the experiments. Sections 3-5 detail the design and results for the experiments on labor market stigmas. Section 6 discusses our results on welfare stigma, and Section 7 concludes.

2 Local Context

Our study takes place in the greater Cairo area of Egypt, a middle-income country with a PPP-adjusted GDP per capita of about \$12,000. In 2016, Egypt faced a 33.4% unemployment rate among workers age 15-24, among the highest of any country (ILO, 2016).

There are many possible supply- and demand-side explanations for Egypt's labor market woes, which predate the political instability of the last decade. We focus on the negative stigmas surrounding available entry-level positions. Unemployed youth may prefer to remain unemployed rather than work in the jobs that are available, perhaps because the jobs are looked down upon in society or professionally unappealing. Anecdotally, policymakers, NGOs, and job seekers in Egypt expressed to us that this is important for understanding the Egyptian labor market. In a study conducted in Jordan, Groh et al. (2015) find strong evidence consistent with this type of behavior.

In our first two experiments, we partnered with the Egypt office of a well-known job matching and training NGO called Education for Employment (EFE). EFE focuses on providing "demand-driven" training by partnering with employers who are looking to hire, and they train people in line with the skills needed for those jobs. The majority of their training is focused on preparing young, college-educated individuals for entry-level service jobs in hotels, restaurants, and retail shops.

We worked with EFE to design different methods to recruit individuals for these training programs, which we outline below. While EFE was successful in filling most of their training classes, doing so was a regular challenge despite providing a highly subsidized (often free) training program. In line with others working in this area, they thought that part of the problem was the stigma associated with working in these entry-level jobs. In Experiment 3, we worked with JobMaster, a human resource company that hosted a job fair for several large companies, and they shared similar concerns about perceived stigmas.

Based on these discussions, we decided to explore four main types of stigma. The first three concern negative feelings about entry-level jobs, while the fourth concerns negative feelings about assistance programs themselves. The first is what we call "social stigma", a sense that entry-level jobs are looked down upon by society, family, and potential marriage partners. This idea came up frequently in our discussions with our Egyptian partners. The second is “professional stigma”, the belief that entry-level jobs are looked down on by future employers, hindering future career progress. The third is “personal stigma”, the internal sense of disappointment associated with performing a job that is not rewarding (e.g., Major et al., 1998). The fourth stigma is “welfare stigma”, the disutility that comes from participating in a program meant for the poor or disadvantaged (Moffitt, 1983).

3 Experiment 1: Job Training Recruitment on Facebook

Experimental Design

We present the three experiments in ascending order of how much information we have about the respondents. Experiment 1 was run on Facebook in late 2018. Facebook is very popular in Egypt, with about 42 million users as of 2020 (Kemp, 2020), and has been used extensively for recruiting trainees by EFE. To explore the importance of stigma, we designed a simple experiment that we implemented on their ad platform. Facebook allows advertisers to run “split tests” which are meant to be randomized experiments of different ads with the ability to compare the performance of ads to each other.

We tested three main ads on Facebook, the exact content of which can be found in the Appendix. The control ad simply informed people about the cotent, length, and format of the training program. We then adjusted the control ad to include additional information about “social stigma” and “professional stigma”. In both cases, we collected testimonials from previous graduates of the training program that described how the types of stigma we thought people would be worried about were in fact not as important as the potential job-seekers may have thought.

In the “social stigma” treatment, we included quotes from past alumni of the training program about how graduating from EFE had led to greater support and respect from their families. For the “professional stigma” treatment, we included examples and quotes from alumni describing promotions and professional growth opportunities they experienced in the years following their graduation and taking of an entry-level position.

Experimenting with Facebook advertising has its benefits and drawbacks. We used their "split test" feature to test the ads against each other. Facebook provides this capability and chooses where and when to serve the ads. They are able to do this with minimal input from the researcher and provide access to large samples for low cost. Our experiment reached 767,768 young people who lived in the greater Cairo area. Individuals were able to click on the ads and sign up for training directly on the Facebook platform. Signing up is our main outcome of interest in this case.

However, Facebook ads are not a very powerful intervention. Most people ignore the ads, reducing the expected impact of the ads and requiring large sample sizes. Other drawbacks include the inability to oversee the randomization itself, and there are also only two binary covariates that are available for those in the sample: gender and age range (18-24 and 25-34).

Unfortunately, even these two covariates are enough to make us question Facebook's randomization algorithm. As Appendix Table A1 shows, the treatments are not balanced by either covariate. Nonetheless, we implement a set of robustness checks that allow us to maintain confidence in our estimated impacts.

Results

We begin our analysis with a straightforward approach, regressing a binary outcome variable (whether the individual signed up for the training on the Facebook platform) on dummies for each treatment, one for "social stigma" and another for "professional stigma". The control group is the excluded category. We present these results in Table 1.

Column 1 provides several notable results. First, the sign-up rate in the control condition is quite low: only 0.12% of individuals served an ad signed up for the training. Despite this, we can still learn from the relative effectiveness of the different ads. We multiply the sign up rate by 100 in Table 1 to make it easier to read the coefficients. Second, we find that both ads that attempt to overcome stigma in fact lead to a *negative* impact on take-up rates. The professional stigma treatment leads to a decrease of 0.032 percentage points, a 26% decrease relative to the control group, and the social stigma treatment leads to a decrease of 0.047, or 39% relative to control. The social stigma effect is significantly larger than the professional stigma effect at the 5% level, suggesting that these two stigmas are distinct phenomena. While we hoped to dispel the negative stigmas with our treatments, we may have just made them more salient instead.[2]

Table 1: Facebook Sign Up Rates (x100) (Experiment 1)

Sample:	Full Sample	Pro & Social Only
	(1)	(2)
Professional Stigma	-0.032 ***	
	(0.009)	
Social Stigma	-0.047 ***	-0.015 **
	(0.009)	(0.008)
Mean of Control Group	0.121	0.087
Number of Observations	767,768	524,979

Notes: This table reports how each Facebook treatment ad affected the proportion of the sample who signed up for the job training program. The dependent variable is multiplied by 100 to make the coefficients easier to read since sign up rates were so low. Robust standard errors in parentheses. Significance * .10; ** .05; *** .01.

The lack of balance in treatments is a concern, so in the Appendix we perform a robustness check similar in spirit to "Lee bounds", where we create balanced groups by dropping individuals who are overrepresented in the treatment groups (Lee, 2009). We identify how many of each type of person is overrepresented in our sample and drop a random subset of them to achieve balance. We implement this procedure 1,000 times, run our regressions with the balanced samples, and plot the treatment coefficients. The results from this robustness check are fully consistent with our original regressions; none of the 1,000 iterations deviates from a finding of a negative and statistically significant impact. The results also hold when we simply include the covariates in our regressions. These exercises make us confident that the stigma treatments indeed caused a decrease in sign-up rates on average.

There are two limitations to our analysis here. First, we have only two binary covariates (gender and age range) and even those are not balanced, so we have little ability to look for heterogeneity in response to the stigma treatments. Second, our outcome is signing up, rather than actually enrolling in the program, because only a

[2] It could be that the depressed take-up is due to the increased length of the ad that individuals see. The professional and social stigma ads are practically the same length, so finding differences between the two stigmas provides evidence that the content of the ad, not just the length, is a primary factor.

handful of individuals enrolled in the program. Nonetheless, showing that stigma is able to affect sign-ups is an important result, since sign-ups necessarily precede enrollment. The other experiments will help us overcome these two limitations.

4 Experiment 2: Street Level Job Training Recruitment

Experimental Design

Our second experiment, implemented from August 2016 to February 2017, used in-person, street-level marketing in different areas of Cairo to recruit for the same training program advertised in Experiment 1. Young adults were approached on the street by a surveyor and asked if they were interested in hearing about a training program being offered for youth interested in finding jobs. If they answered yes, basic eligibility information was collected. If the person was eligible for the training program (as defined by the NGO), more information was collected and they received a randomized recruitment pitch from the surveyor.[3]

As in Experiment 1, we gave pitches aimed at professional and social stigma. The control group was given information about the program's purpose, location, duration, and recent outcomes (the average incomes of individuals who graduated 1 year and 5 years ago). Those in the stigma treatment groups got the same information, plus text that was almost identical to Experiment 1 (exact wording can be found in the Appendix). We also added a third stigma treatment arm focused on "personal stigma", where we included statistics from our alumni survey about job satisfaction as well as quotes from alumni about the enjoyment of their initial job placements.

After hearing the pitch, individuals were invited to sign up for the training on the spot. Conditional on agreeing to apply, they were then asked more detailed questions related to their prior work history and family background.

The street-level recruitment strategy has several advantages over the Facebook recruitment, but also its own limitations. It is much more intense than online ads, allowing for what we expected to be a stronger treatment effect. Speaking to peo-

[3]Eligibility was determined by asking if the respondent is unemployed or underemployed; how old they are; their educational attainment; whether they attended public or private school; and their military status.

ple in person also allows for better screening of individuals and more data collection about their backgrounds. We were particularly interested in collecting data on income, because we thought stigma concerns might be most relevant for those of higher socioeconomic status (SES). To get a proxy for SES, we included a question before the information pitches about the type of transport people primarily take. We classify individuals who take private car or mobility on demand services (e.g., Uber) as "relatively rich". The chief limitation of face-to-face interactions relative to online ads is that the sample is necessarily much smaller (2,900 individuals).

As with Experiment 1, our outcome of interest is applying for the program. Application rates are orders of magnitude larger, with the control group applying for the program 43% of the time. Unfortunately, only a handful of applicants ended up participating in the training, leading to insufficient power to detect effects on enrollment. Experiment 3 will overcome this limitation.

Average Results

Table 2 reports the results from Experiment 2. Panel A reports the average impacts. Across all three stigma treatments, we find statistically insignificant effects on take-up, with negative point estimates for social and professional stigma.[4]

Using Machine Learning to Test for Heterogeneity

However, the average results mask considerable heterogeneity, which we present in Panels B and C. We assess heterogeneity in two different ways. First, we utilize methods from Chernozhukov et al. (2020), which provides a strategy for detecting heterogeneity in an "agnostic" fashion. This strategy utilizes generic machine learning methods and sample-splitting cross-validation techniques, using baseline covariates to estimate two models that predict the outcome of interest (in our case applying for the program) depending on whether an individual is in the treatment group or control group. It splits the sample so that it can generate a model in the "training set" and produces

[4]We also cross-randomized the price of the program from a small fee of 200 EGP, or about $25, to an incentive payment of about $12.50 (the actual cost to provide the program was around 4500 EGP). We found that application rates decrease with price but do not increase with the incentive. We control for this cross randomization in our analysis.

Table 2: Job Training Application Rates (Experiment 2)

Panel A: Average Treatment Effect	Social (1)	Professional (2)	Personal (3)
Treatment Indicator	-0.019	-0.029	0.006
	(0.026)	(0.026)	(0.025)
Mean of Control Group	0.331	0.331	0.331
Number of Observations	1460	1470	1464
Panel B: Heterogeneous Treatment Effects by ML Group			
Treatment Effect for Top Group	0.054	0.074	0.073
	(0.058)	(0.056)	(0.056)
Treatment Effect for Bottom Group	-0.155 ***	-0.125 **	0.047
	(0.056)	(0.058)	(0.058)
P-Value for difference between groups	0.009	0.015	0.756
Number of Observations	1460	1470	1464
Panel C: Compositional Difference in Applicants, Treatment vs. Control			
Lasso-Based Index	0.174 **	0.209 ***	0.148 **
	(0.083)	(0.079)	(0.075)
Individual Characteristics:			
Age	0.125	0.235	-0.097
	(0.198)	(0.205)	(0.196)
Male	0.033	0.046	-0.025
	(0.036)	(0.037)	(0.037)
Rich	0.048	0.087 ***	0.051 *
	(0.031)	(0.032)	(0.030)
Currently Working	0.045	0.070 **	0.034
	(0.035)	(0.035)	(0.034)
University Degree	-0.023	0.002	-0.045
	(0.035)	(0.036)	(0.035)
Number of Observations	611	611	631

Notes: Column 1 reports the results for the "Social Stigma" treatment, Columns 2 & 3 report results for the "Professional Stigma" & "Personal Stigma" treatments respectively. Panel A reports how the treatment messages affected the proportion of indiviudals who applied for the program on average. Panel B reports the effects of the treatment on the individuals in the top & bottom "Individual Treatment Effect" groups as assigned by the methods from Chernozhukov et al (2020). Panel B also reports the p-value for a test of equality of coefficients for the top and bottom groups. Panel C compares the average characteristics of *applicants* in the treatment and control groups. The top row shows how the applicants differ on an index that was produced by running a lasso regression of the "Individual Treatment Effects" on all pairwise combinations of our baseline data. The following rows show differences between applicants in treatment and control on a subset of baseline variables. Robust standard errors in parentheses. Significance * .10; ** .05; *** .01.

valid predictions in the "testing set". It then takes the difference between these models to be the "individual treatment effect" (ITE). The sample is then sorted by their ITEs and split into 5 ordered groups, which are used to estimate the "Group Average Treatment Effect". In essence, this strategy identifies which people have the largest predicted response to treatment and which have the lowest. We explain the details of our implementation of this method in the Appendix.[5]

In Panel B of Table 2, we report the estimated treatment effects for those in the highest estimated ITE group and those with the lowest estimated ITEs. We first consider the "social stigma" treatment, which was the most impactful treatment in Experiment 1. Those in the highest ITE group increase their application rates slightly in response to this treatment, while those in the lowest ITE group *decrease* their application rates by a highly significant 15.5 percentage points. The difference between the treatment effects for these top and bottom groups is significant at the 1% level ($p = 0.009$). We take this as strong evidence of heterogeneous treatment effects. We repeat the analysis for professional stigma and again find a large and statistically significant difference in the effects on the top and bottom group ($p = 0.015$). On the other hand, we do not find any evidence of heterogeneity in the effects of the personal stigma treatment.

We utilize a second strategy to assess heterogeneity in Panel C, where we perform a compositional analysis. In particular, we analyze whether the characteristics of applicants in the treatment group differ from those in the control group – that is, whether the stigma treatments changed the composition of applicants. Even if stigma has no average effect on take-up (as we found in Panel A), heterogeneous effects could alter the composition of who applies, because the treatments may work differently on different subgroups in our sample.

We first implement a Lasso regression of the estimated individual treatment effect (ITE) on our baseline characteristics and their pairwise interactions. This produces an index of characteristics that predicts who will respond more to treatment. We then

[5]We differ from the method outlined in the Chernozhukov et al. (2020) paper in two ways in order to maximize statistical power, which we explain in more detail in the Appendix. First, we produce groups using the full sample by taking the median ITE across all simulations. Second, we compare the coefficients of the top and bottom groups, as opposed to assuming a linear relationship for the individual treatment effect.

compare the value of that index for applicants in the control group and applicants in the treatment group and find that in the social stigma treatment, applicants from the treatment group score 0.174 standard deviations higher on this index than applicants from the control group, with the difference significant at the 5% level.[6] This proves that there was a meaningful impact of the stigma treatment; it changed the composition of who applied to the program and who did not.

The Lasso-based index does not provide a simple interpretation, as is common in machine learning analysis (Mullainathan and Spiess, 2017). In Panel C, we test how individual characteristics of applicants differ between the treatment and control groups in a way that helps us better interpret which groups are being affected by the stigma treatment. Applicants from the treatment group are older, more likely to be male, richer, and more likely to be currently working, though none of these is significant for social stigma.

We then repeat this analysis for the professional and personal stigma treatments. We find large differences in the composition of applicants in the professional stigma treatment, relative to control. The differences are similar to those in the social stigma treatment, being richer and more likely to be working. For social stigma, despite not finding statistically significant differences in the treatment effects of those in the highest and lowest ITE groups (Panel B), we do find that the treatment significantly altered the composition of applicants (Panel C). This shows that this treatment is nonetheless affecting different people differently. Individuals who apply for the program from the treatment groups score higher on the Lasso-based index relative to applicants from the control group. In the personal stigma treatment, treatment applicants are again richer and more likely to be currently working than the control group applicants.

These results show that stigmas surrounding entry-level jobs are an important factor in application behavior of job-seekers, but the effects are heterogeneous. Stigma does not seem to depress overall take-up, but it changes the composition of who applies. Depending on who the program wants to target, this may or may not be ideal, and it

[6]This result is not mechanical. The machine learning methods that predict the individual treatment effects (ITEs) utilize split-sample techniques which ensure that the ITEs estimated for each person are "honest", i.e. estimated using data on other applicants and not their own characteristics. Without the split-sample validation we could be worried that we are using data on one individual's behavior to predict their own behavior which would mechanically lead to success.

could alter the effectiveness of the program (Card et al., 2018).

5 Experiment 3: Door-to-Door Recruitment for a Job Fair

Experimental Design

In Experiment 3, our focus changes slightly. Instead of recruiting individuals to attend a multi-week job training program, we implemented a door-to-door information campaign in December 2019 to encourage people to attend a upcoming free job fair. The job fair was focused on the same types of entry-level service sector jobs as the training program, but should have higher participation rates because a one-time event requires less commitment.

In this experiment, surveyors went from apartment to apartment, asking if there was anyone in the household who was looking for a job. If yes, they would check to see if that individual was in the same age range as the training (18 to 35). They would then collect some basic demographic information and read a randomized informational message about the job fair. To decrease the potential for information spillovers, the randomization was implemented at the building level. We cluster our results for this experiment by building.

A key focus of Experiment 3 is trying to distinguish between making a negative stigma more salient and actually dispelling that stigma. In this experiment, the control group received a message that provided information about the time and location of the job fair and the firms and types of jobs that would be available there. Treatment 1 was meant to make social stigma salient: we included the statement, "Although some people might think these types of jobs might be looked down on in society, it's important to start somewhere." Treatment 2 was meant to bring up the stigma *and* dispel it, replacing "it's important to to start somewhere" with, "actually people in these types of jobs report that their families respect and encourage them more than before they had a job." We then included testimonials as in Experiments 1 and 2, which can be found in the Appendix.

Although it has a smaller sample than the previous experiments, Experiment 3 provides two main benefits. First, our outcome is actual attendance of the job fair rather than just an application. About 6% of the control group showed up to the job fair, allowing us to test if stigma affects actual labor market behavior. While both attendance and signing up for training are merely first steps to real employment, showing up to a fair takes more time and effort than simply signing up for something. Second, we know more about the recruits than we did in the other experiment, including their education, work status, and job aspirations, which provides the machine learning algorithms more data to use for prediction.

Average Results

Table 3 reports results with attendance as the outcome. Panel A shows that, again, merely alluding to stigma decreases attendance by 1.4 percentage points. Attempting to dispel it is not effective, also decreasing attendance by 2.7 percentage points. While statistically insignificant, these effects are large in relative terms, as only 5.9% of the control group attended the fair. Our results could be an example of "ironic rebound", where the mere mention of stigma increases its salience, even if the intent is to dispel it (Wegner et al., 1987).

Using Machine Learning to Test for Heterogeneity

As in Experiment 2, we implement the machine learning techniques from Chernozhukov et al. (2020) to detect whether there are heterogeneous treatment effects. In Panel B, we report the estimated effects on the individuals who were predicted to have the highest individual treatment effects (ITEs) and those predicted to have the lowest ITEs. For salient stigma, we again find evidence of heterogeneous treatment effects: the treatment has a strong negative effect on some people and no effect on others. This difference is statistically significant ($p = 0.022$). On the other hand, the dispelling stigma treatment does not lead to as much of a negative effect on the bottom end, and so we do not see significant heterogeneity in responses.

Panel C considers whether the stigma treatments affected the composition of program participants. We again generate an index based on a Lasso regression of the estimated ITE on all baseline characteristics and their pairwise interactions. We find that applicants in the salient treatment arm are more likely to score higher on this index relative to applicants in the control group. These individuals are much older

(2.8 years), almost twice as likely to be currently working, and slightly richer. While only age is significant due to small sample sizes here, the directions of the effects are consistent with those from Experiment 2. In both experiments, the stigma treatments are leading to a group of applicants/participants who differ in these same ways.

6 Testing for Welfare Stigma

The most familiar type of stigma to many economists is "welfare stigma". While the stigmas we have focused on so far surround the entry-level jobs, welfare stigma is the stigma associated with participating in a program intended for the poor or less fortunate (e.g., Moffitt, 1983). While welfare stigma is a common feature of discussions about take-up of social programs, evidence of its importance is scarce (Currie, 2006). In the only two experimental treatments we are aware of, Bhargava and Manoli (2015) find little evidence of welfare stigma for EITC take-up in the US, while Friedrichsen et al. (2018) find evidence for it in a lab setting. We are not aware of any evidence on welfare stigma in developing countries.

In Experiment 2, the street-level recruitment for job training, we also tested for the importance of welfare stigma. After telling recruits the price they would pay for the training program, we randomly told some of them that the "true cost" of the program is usually higher but has been reduced through donations from organizations. Within those who get this information, a random subset was also told that the price had been reduced "to help those in financial hardship". This is the welfare stigma treatment. If welfare stigma is relevant here, those getting this treatment should have lower application rates than those who get only the true cost information.

Attending a multi-week training program is likely more visible than receiving government benefits, so we might expect welfare stigma to be especially relevant here. On the other hand, job training programs are common in Egypt, so there may be little social stigma associated with taking part in one. This specific training program is not widely known, which might also reduce the stigma. Even if there is not this type of social stigma, though, there could be a personal welfare stigma associated with taking any assistance intended for the poor.

Table 4 shows the results. Overall, we find no effect on take-up rates in Panel

Table 3: Job Fair Attendance Rates (Experiment 3)

Panel A: Average Treatment Effect	Salient (1)		Dispelling (2)	
Treatment Indicator	-0.014		-0.027	
	(0.015)		(0.017)	
Mean of Control Group	0.059		0.059	
Number of Observations	768		746	
Panel B: Heterogeneous Treatment Effects by ML Group				
Treatment Effect for Top Group	0.006		-0.016	
	(0.316)		(0.027)	
Treatment Effect for Bottom Group	-0.120	***	-0.044	
	(0.042)		(0.043)	
P-Value for difference between groups	0.022		0.585	
Number of Observations	768		746	
Panel C: Compositional Difference in Applicants, Treatment vs. Control				
Lasso-Based Index	0.914	**	-0.089	
	(0.411)		(0.617)	
Individual Characteristics:				
Age	2.792	**	1.375	
	(1.254)		(1.711)	
Male	0		0.25	*
	(0.183)		(0.142)	
Rich	0.028		0.000	
	(0.092)		(0.098)	
Currently Working	0.181		0.208	
	(0.133)		(0.191)	
University Degree	0.042		-0.042	
	(0.129)		(0.160)	
Number of Observations	42		36	

Notes: Column 1 reports the results for the "Salient Stigma" treatment, Column 2 reports results for the Dispelling Stigma" treatments. Panel A reports how the treatment messages affected the proportion of indiviudals who attended the job fair on average. Panel B reports the effects of the treatment on the individuals in the top & bottom "Individual Treatment Effect" groups as assigned by the methods from Chernozhukov et al (2020). Panel B also reports the p-value for a test of equality of coefficients for the top and bottom groups. Panel C compares the average characteristics of *attendees* in the treatment and control groups. The top row shows how the attendees differ on an index that was produced by running a lasso regression of the "Individual Treatment Effects" on all pairwise combinations of our baseline data. The following rows show differences between applicants in treatment and control on a subset of baseline variables. Robust standard errors in parentheses. Significance * .10; ** .05; *** .01.

A, and we can reject any large negative effect. In Panels B and C, we again use our two methods to assess any heterogeneity in treatment effects. In contrast to the results on the stigmas surrounding entry-level jobs, we find little evidence that welfare stigma is important for any group. Panel B uses the machine learning strategy from Chernozhukov et al. (2020) and finds no significant difference in treatment effects for the top and bottom groups, and no significant effect for any group. In Panel C, we find a marginally significant difference in composition of applicants by the Lasso-based index, but there are no differences on any observable characteristics. One might have expected to find a heterogeneous effect by wealth when looking at welfare stigma, but we do not find this.

Table 4: Testing for Welfare Stigma (Experiment 2)

Panel A: Average Treatment Effect		**Panel B: Heterogeneous Treatment Effects by ML Group**	
	Welfare Stigma		Welfare Stigma
	(1)		(2)
Treatment Indicator	0.003	Treatment Effect for Top Group	0.043
	(0.022)		(0.050)
		Treatment Effect for Bottom Group	-0.026
			(0.049)
Mean of Control Group	0.426	P-Value for difference between groups	0.315
Number of Observations	1949	Number of Observations	1949

Panel C: Compositional Difference in Applicants, Treatment vs. Control (N=834)

Lasso-Based Index	Age	Male	Rich	Currently Working	University Degree
(3)	(4)	(5)	(6)	(7)	(8)
0.120 *	0.143	-0.031	-0.002	-0.023	0.005
(0.065)	(0.173)	(0.031)	(0.028)	(0.030)	(0.029)

Notes: Column 1 reports how the welfare treatment messages affected the proportion of indiviudals who applied for job training. The treatment group was told that the job training was "subsidized for the poor", while the control group in this table are those who were told that training was only "subsidized". Panel B reports the effects of the treatment on the individuals in the top & bottom "Individual Treatment Effect" groups as assigned by the methods from Chernozhukov et al (2020). Panel B also reports the p-value for a test of equality of coefficients for the top and bottom groups. Panel C compares the average characteristics of *applicants* in the treatment and control groups. Column 3 shows how the applicants differ on an index that was produced by running a lasso regression of the "Individual Treatment Effects" on all pairwise combinations of our baseline data. The following columns show differences between applicants in treatment and control on a subset of baseline variables. Robust standard errors in parentheses. Significance * .10; ** .05; *** .01.

Overall, we find much stronger evidence for the importance of stigmas surrounding entry-level jobs (at least for some groups of people) than for the importance of any welfare stigma related to the job training program. This could help explain why the evidence for stigma with regard to social programs is so thin (e.g., Currie, 2006) while the evidence for stigma and social image affecting how people behave and invest in their careers is stronger (e.g., Bursztyn and Jensen, 2015).

7 Discussion and Conclusion

In three randomized experiments in Egypt, we provide novel evidence on the impacts of several types of stigma on the take-up and composition of labor market assistance programs. Negative stigma associated with the prospects and social image of entry-level jobs is clearly an important factor in labor market decisions.

Attempts to counteract the stigmas associated with entry-level jobs were unsuccessful on average, but this masks strong heterogeneity. Agnostic machine learning techniques show evidence that different people respond very differently to our treatments. Some respond positively, while others have large negative effects. This means that attempts to overcome stigma have significant effects on the composition of who participates in a program, even if the overall effect on take-up is small. In our experiments, the stigma treatments deliver an applicant pool that is older, richer, and more likely to be currently working than we get from the control group. This could affect the returns to these programs.

On the other hand, we find no evidence that welfare stigma affects take-up or composition of these programs. There may be stronger stigmas related to programs' job outcomes than to the social programs themselves.

For policymakers, our results showcase that messaging around programs is of first-order importance. Stigma may not have large effects on the level of take-up, but it can significantly change the composition of a program, and potentially its effectiveness. Additional research on stigma is needed. Careful testing of recruitment messages can help inform strategies for targeting the groups best suited for the program.

References

Acevedo, P., G. Cruces, P. Gertler, and S. Martinez (2020): "How Vocational Education Made Females Better Off but Left Males Behind," Labour Economics, 65, 101824.

Attanasio, O., A. Guarin, C. Medina, and C. Meghir (2017): "Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow Up," American Economic Journal: Applied Economics, 9, 131–143.

Attanasio, O., A. Kugler, and C. Meghir (2011): "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial," American Economic Journal: Applied Economics, 3, 188–220.

Bandiera, O., R. Burgess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman (2017): "Labor markets and poverty in village economies," The Quarterly Journal of Economics, 132, 811–870.

Besley, T. and S. Coate (1995): "Group lending, repayment incentives and social collateral," Journal of Development Economics, 46, 1–18.

Bhargava, S. and D. Manoli (2015): "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment," American Economic Review, 105, 3489–3529.

Bursztyn, L., M. Callen, B. Ferman, S. Gulzar, A. Hasanain, and N. Yuchtman (2016): "Political identity: Experimental evidence on anti-Americanism in Pakistan," Journal of the European Economic Association.

Bursztyn, L., A. Gonzalez, and D. Yanagizawa-Drott (2020): "Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia," American Economic Review, 110, 2997–3029.

Bursztyn, L. and R. Jensen (2015): "How Does Peer Pressure Affects Educational Investments?" Quarterly Journal of Economics, 130, 1329–1367.

——— (2017): "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure," Annual Review of Economics, 9, 131–153.

Card, D., J. Kluve, and A. Weber (2018): "What works? A meta analysis of recent active labor market program evaluations," Journal of the European Economic Association, 16, 894–931.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2020): "Generic machine learning inference on heterogeneous treatment effects in randomized experiments," National Bureau of Economic Research Working Paper 24678.

Cruces, G., R. Perez-Truglia, and M. Tetaz (2013): "Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment," Journal of Public Economics, 98, 100–112.

Currie, J. (2003): "U.S. Food and Nutrition Programs," in Means-Tested Welfare Programs in the United States, ed. by R. Moffitt, University of Chicago Press for NBER.

——— (2006): "The Take-Up of Social Benefits," in Poverty, the Distribution of Income, and Public Policy, ed. by D. C. Alan Auerbach and J. Quigley, Russell Sage.

Friedrichsen, J. and D. Engelmann (2018): "Who cares about social image?" European Economic Review, 110, 61–77.

Friedrichsen, J., T. König, and R. Schmacker (2018): "Social image concerns and welfare take-up," Journal of Public Economics, 168, 174–192.

Ghosal, S., S. Jana, A. Mani, S. Mitra, and S. Roy (2020): "Sex Workers, Stigma, and Self-Image: Evidence from Kolkata Brothels," Review of Economics and Statistics, 1–45.

Groes, F., P. Kircher, and I. Manovskii (2015): "The U-shapes of occupational mobility," The Review of Economic Studies, 82, 659–692.

GROH, M., N. KRISHNAN, D. MCKENZIE, AND T. VISHWANATH (2016): "Do wage subsidies provide a stepping-stone to employment for recent college graduates? Evidence from a randomized experiment in Jordan," Review of Economics and Statistics, 98, 488–502.

GROH, M., D. MCKENZIE, N. SHAMMOUT, AND T. VISHWANATH (2015): "Testing the importance of search frictions and matching through a randomized experiment in Jordan," IZA Journal of Labor Economics, 4, 7.

HECKMAN, J. J. AND J. A. SMITH (2004): "The determinants of participation in a social program: Evidence from a prototypical job training program," Journal of Labor Economics, 22, 243–298.

HIRSHLEIFER, S., D. MCKENZIE, R. ALMEIDA, AND C. RIDAO-CANO (2016): "The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey," Economic Journal, 127, 2115–2146.

ILO (2016): "Employent by Sex and Age – ILO Modelled Estimates," ILOSTAT. Accessed 02-14-2020. `https://ilostat.ilo.org/data`.

KEMP, S. (2020): "Digital 2020: Egypt," Retrieved from https://https://datareportal.com/reports/digital-2020-egypt on 2020-06-19.

KLUVE, J., S. PUERTO, D. ROBALINO, J. M. ROMERO, F. ROTHER, J. STÖTERAU, F. WEIDENKAFF, AND M. WITTE (2019): "Do youth employment programs improve labor market outcomes? A quantitative review," World Development, 114, 237–253.

LEE, D. S. (2009): "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," The Review of Economic Studies, 76, 1071–1102.

MAJOR, B., S. SPENCER, T. SCHMADER, C. WOLFE, AND J. CROCKER (1998): "Coping with negative stereotypes about intellectual performance: The role of psychological disengagement," Personality and Social Psychology Bulletin, 24, 34–50.

MANCHESTER, C. F. AND K. J. MUMFORD (2010): "Welfare stigma due to public disapproval," University of Minnesota and Purdue University, unpublished working paper.

McKenzie, D. (2017): "How effective are active labor market policies in developing countries? a critical review of recent evidence," The World Bank Research Observer, 32, 127–154.

Moffitt, R. (1983): "An economic model of welfare stigma," The American Economic Review, 73, 1023–1035.

Mullainathan, S. and J. Spiess (2017): "Machine learning: an applied econometric approach," Journal of Economic Perspectives, 31, 87–106.

Remler, D. K. and S. A. Glied (2003): "What other programs can teach us: Increasing participation in health insurance programs," American Journal of Public Health, 93, 67–74.

Rinehart, C. S. and J. W. McGuire (2017): "Obstacles to Takeup: Ecuador's Conditional Cash Transfer Program, The Bono de Desarrollo Humano," World Development, 97, 165–177.

Schofield, T. P., N. Haslam, and P. Butterworth (2019): "The persistence of welfare stigma: Does the passing of time and subsequent employment moderate the negative perceptions associated with unemployment benefit receipt?" Journal of Applied Social Psychology, 49, 563–574.

Sicherman, N. and O. Galor (1990): "A theory of career mobility," Journal of political economy, 98, 169–192.

Stuber, J. and M. Schlesinger (2006): "Sources of stigma for means-tested government programs," Social Science & Medicine, 63, 933–945.

Wegner, D. M., D. J. Schneider, S. R. Carter, and T. L. White (1987): "Psychological Effects of Thought Suppression," Journal of Personality and Social Psychology, 53.

World Bank Group (2018): World development report 2019: The changing nature of work, World Bank.

Appendix: Scripts used in Each Experiment

Experiment 1: Facebook

- **Control:** "Are you looking for a job but you don't know where to start? EFE will help you take the first step in your professional career through the "Job Placement Training Program". EFE works with private sector companies to ensure that the skills that you will acquire from the program are the skills that the job market really needs and we help you obtain job interviews with well-known companies before you graduate. We are starting a JPTP program and the program is grant supported by and so you can get it at no cost. You must commit to the full scholarship period (every day from Sunday to Thursday from 9 am to 5:30 pm for 3 weeks). Apply!"

- **Professional Stigma:** "Are you looking for a job but you don't know where to start? EFE will help you take the first step in your professional career through the "Job Placement Training Program". EFE works with private sector companies to ensure that the skills that you will acquire from the program are the skills that the job market really needs and we help you obtain job interviews with well-known companies before you graduate. Note that although some people might think that these types of jobs might be a professional dead-end, graduates of the program who started in these jobs often end up climbing the professional ladder to become managers and directors. Overall there is a high rate of professional development amongst graduates of EFE who have taken these jobs. For example, one EFE alumnus started as a content associate and 5 years later he is currently a senior content supervisor managing a team of over 60 employees. We also interviewed some recent alumni who said "I definitely felt like there was scope to grow in my first job", and "There was definitely room to grow professionally, 100%. You must commit to the full scholarship period (every day from Sunday to Thursday from 9 am to 5:30 pm for 3 weeks). Apply!"

- **Social Stigma:** "Are you looking for a job but you don't know where to start? EFE will help you take the first step in your professional career through the "Job Placement Training Program". EFE works with private sector companies to ensure that the skills that you will acquire from the program are the skills that the job market really needs and we help you obtain job interviews with well-known companies before you graduate. We are starting a JPTP program and the program is grant supported by and so you can get it at no cost. Note that although some people might think that these types of jobs might be looked down on in society, graduates of EFE who have taken these jobs report that their families and communities hold them in higher regard. For example, one alumnus recently said about his experience, "[My father] now supports me and encourages me to excel a lot more than he did in the past." Another alumnus said, "My parents have always been very supportive of me, but they are definitely proud of me now. We are starting a JPTP program and the program is grant supported by and so you can get it at no cost. You must commit to the full scholarship period (every day from Sunday to Thursday from 9 am to 5:30 pm for 3 weeks). Apply!"

Experiment 2: Job Training

- **Control:** "I am from Education for Employment| Egypt. We are an organization that specializes in youth training and employment and focuses on improving the skills of graduates to support them in securing job opportunities through developing some of their skills such as presentation, communication, CV writing skills, computer skills, and the English language skills needed by the labor market.

  The training program normally takes about 3-4 weeks, and is located at [NGO Address]. It takes place six days a week from 9am-5.30pm, and is usually conducted in classes of 25 or so students, who all work together on a variety of topics to help them learn more about the skills that are needed in the labor market. The training takes on an interactive and practical approach and ensures that students learn how to utilize those skills to turn them into fruitful employment opportunities after graduation. We provide certificates of completion and help you find jobs after you finish the program. We also provide access to a large network of over 2000 graduates with similar profiles, and a variety of ongoing professional development courses after graduation.

We ensure that all programs are market-driven and based on the needs of the local labor market. When implementing programs, we establish partnerships with private sector employers that have a demand for new high-quality employees. We are funded by a variety of sources and our phone number is [NGO Phone Number].

This training program aims to help individuals find employment opportunities that will help them grow professionally in the future. In the past, our graduates have gotten jobs like waiters, retailers, marketers, sales associates, call center agents, and e-commerce associates, etc. Average starting salaries for employed graduates are 1450 LE per month, and after 3 years the average employed person is making about 3400 LE per month."

- **Professional Stigma:** Control Pitch + "Note that although some people might think that these types of jobs might be a professional dead-end, graduates of the program who started in these jobs often end up climbing the professional ladder to become managers and directors. Overall there is a high rate of professional development amongst graduates of EFE who have taken these jobs. For example, one EFE alumnus started as a content associate and 5 years later he is currently a senior content supervisor managing a team of over 60 employees. We also interviewed some recent alumni who said "I definitely felt like there was scope to grow in my first job", and "There was definitely room to grow professionally, 100%."

- **Social Stigma:** Control Pitch + "Note that although some people might think that these types of jobs might be looked down on in society, graduates of EFE who have taken these jobs report that their families and communities hold them in higher regard. For example, one alumnus recently said about his experience, "[My father] now supports me and encourages me to excel a lot more than he did in the past." Another alumnus said, "My parents have always been very supportive of me, but they are definitely proud of me now."

- **Personal Stigma:** Control Pitch + "Note that although some people might think that these types of jobs are not very enjoyable, there is actually high satisfaction among graduates of EFE who have taken these jobs. For example, one alumnus recently said, 'I definitely enjoyed my first job because the workplace was very positive, and I got to know new people.' And our records indicate that about 80% of EFE alumni stayed in their first job for more than 1 year."

- **True Cost:** Control Pitch + Stigma Pitch + "The true cost of the program is usually around 4000 LE, but many organizations have donated to EFE Egypt so that we can provide this at a much lower cost."

- **Welfare Stigma:** Control Pitch + Stigma Pitch + "The true cost of the program is usually around 4000 LE, but many organizations have donated to EFE Egypt so that we can provide this at a much lower cost. These funds are meant to help those in financial hardship."

Experiment 3: Job Fair

- **Info Pitch (Control):** "Job Master is holding a job fair on December 14th at Ard El-Maared. A job fair is a group of companies offering different jobs for job seekers to apply on the spot. Many companies will be attending such as [5 well known employers] and others. They are trying to hire nearly 400 jobs including entry level positions in sales, drivers, security, electrical/mechanical maintenance technicians, warehouse workers, etc. Salaries range between 1900 and 4000, and could reach up to 8000 for certain specializations and supervisory functions. Attendance is free."

- **Salient Stigma:** "Job Master is holding a job fair on December 14th at Ard El-Maared. A job fair is a group of companies offering different jobs for job seekers to apply on the spot. Many companies will be attending such as [5 well known employers] and others. They are trying to hire nearly 400 jobs including entry level positions in sales, drivers, security, electrical/mechanical maintenance technicians, warehouse workers, etc. Salaries range between 1900 and 4000, and could reach up to 8000 for certain specializations and supervisory functions. Attendance is free. Although some people might think some of these entry level jobs are looked down on in society, it's important to start somewhere."

- **Dispelling Stigma:** "Job Master is holding a job fair on December 14th at Ard El-Maared. A job fair is a group of companies offering different jobs for job seekers to apply on the spot. Many companies will be attending such as [5 well known employers] and others. They are trying to hire nearly 400 jobs including entry level positions in sales, drivers, security, electrical/mechanical maintenance technicians, warehouse workers, etc. Salaries range between 1900 and 4000, and could reach up to 8000 for certain specializations and supervisory functions. Attendance is free. Although some people might think some of these entry level jobs are looked down on in society, people in these types of jobs report that their families respect and encourage them more than before they had a job. For example, one person we recently talked to who took an entry-level position said about his experience, "[My father] now supports me and encourages me to excel a lot more than he did in the past." Another person we talked to said, "My parents have always been very supportive of me, but they are definitely proud of me now."

Appendix:Balance Table

Table A1: Balance Table

Panel A: Facebook and Job Fair Experiments

	FB Experiment				Job Fair Experiment			
	Control	Professional Stigma	Social Stigma	Combined Stigma	Control	Salient Stigma	Dispelling Stigma	Combined Stigma
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.53	-0.03***	-0.01***	-0.05***	0.37	0.03	0.05	0.04
	{0.50}	(0.00)	(0.00)	(0.00)	{0.03}	(0.03)	(0.04)	(0.03)
Old/Age	0.28	0.03***	0.01***	0.04***	25.42	-0.48	-0.42	-0.45
	{0.02}	(0.001)	(0.001)	(0.001)	{0.34}	(0.43)	(0.46)	(0.40)
University					0.20	0.02	0.00	0.01
					{0.02}	(0.03)	(0.03)	(0.03)
Rich					0.10	0.00	0.02	0.01
					{0.02}	(0.02)	(0.03)	(0.02)
Working					0.45	0.00	-0.01	0.00
					{0.03}	(0.04)	(0.04)	(0.03)
P-value		0.00	0.00	0.00		0.47	0.522	0.445

Panel B: Street Level Experiment

	Control	Personal Stigma	Professional Stigma	Social Stigma	Combined Stigma	Control	Welfare Stigma	True Cost
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.64	-0.02	0.03	0.04	0.02	0.63	0.02	0.04*
	{0.48}	(0.02)	(0.02)	(0.02)	(0.02)	{0.48}	(0.02)	(0.02)
Age	24.70	0.13	0.18	0.20	0.17	24.70	0.07	0.01
	{2.43}	(0.13)	(0.13)	(0.13)	(0.10)	{2.45}	(0.09)	(0.09)
University	0.66	-0.03	-0.03	-0.03	-0.03	0.62	0.02	0.02
	{0.48}	(0.02)	(0.02)	(0.02)	(0.02)	{0.49}	(0.02)	(0.02)
Rich	0.19	0.01	0.037	0.00	0.01	0.22	-0.04**	-0.01
	{0.39}	(0.02)	(0.02)	(0.02)	(0.02)	{0.41}	(0.01)	(0.02)
Working	0.34	0.03	0.02	0.04	0.03	0.34	0.02	0.02
	{0.47}	(0.02)	(0.02)	(0.02)	(0.02)	{0.47}	(0.02)	(0.02)
P-Value		0.630	0.552	0.435	0.648		0.0747	0.217

Notes: This table reports how baseline characteristics differ by group. Columns 1-4 of Panel A report differences for Experiment 2 (Facebook Ads), and Columns 5-8 report differences for Experiment 3 (Job Fair Recruitment). Panel B reports differences for Experiment 3 (Street Level Recruitment). The tables also report p-values for the joint test of all reported baseline covariates on treatment assignment relative to control. Standard deviations for the control group in brackets. Robust standard errors in parentheses. Significance * .10; ** .05; *** .01

Appendix: Robustness Tests for Experiment 1

The Facebook platform provides the opportunity to run "split tests" which are intended to be randomized experiments that test which ad is most effective in inducing individuals to sign up. In checking the two baseline covariates that Facebook provides (a binary for age and a binary for gender) we found that they were unbalanced, which brings into question whether or not we can trust the results of the experiment.

As a robustness check we implement a procedure that creates a balanced sample by randomly dropping "excess" observations across treatment arms by covariate. For example, if the control arm is perfectly balanced by gender and age (50% female, 50% "young"), while the social stigma arm has additional women (e.g. 52% male), then we will be able to determine how many "excess" men there are to achieve balance. In this example it would be 4 percent of the total sample, so that after we remove 4 percentage points of men we get down to 48% men and 48% women. We then randomly choose this proportion of men and drop them from the sample, and implement the same procedure for each treatment on both gender and age. This produces a "manually balanced" sample. We then implement this procedure using 1,000 different random seeds and plot the regression coefficients for treatment in each sample. This produces a set of balanced samples whose results we compare with our primary regression coefficients. Appendix Figures A1 & A2 show that throughout the 1,000 iterations we consistently find a negative and statistically significant impact of the treatments on sign up rates, allowing us to be confident that the results are not being driven by a lack of covariate balance.

Figure A1: Robustness of Professional Stigma Effects
to Alternate Balanced Samples

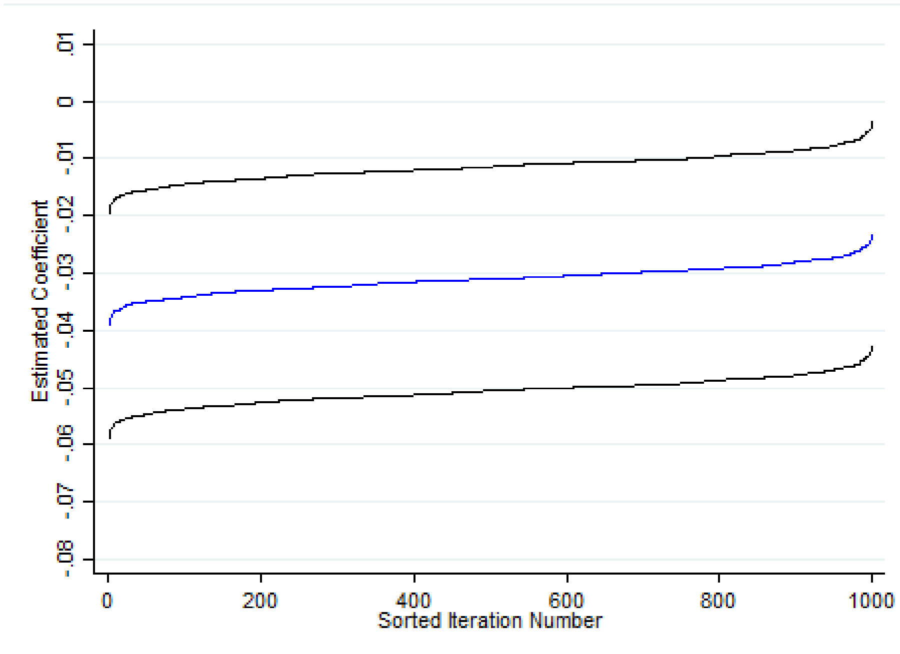
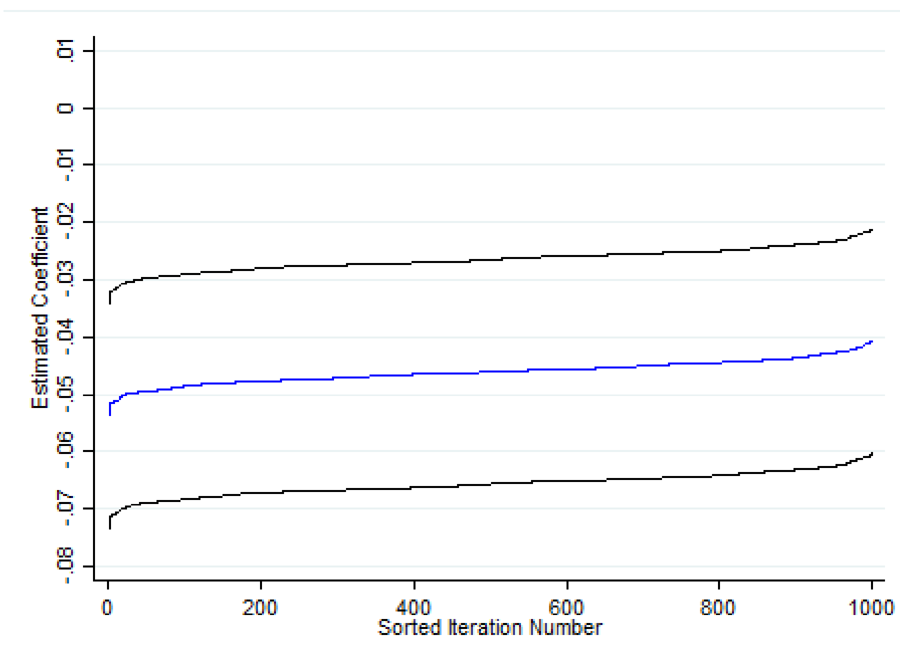


Figure A2: Robustness of Social Stigma Effects
to Alternate Balanced Samples



Appendix: Implementation of Machine Learning for Heterogeneous Treatment Effects

The methods we use for assessing heterogeneity in treatment effects from randomized experiments are taken from Chernozhukov et al. (2020). The method they put forth can be summarized in the following steps:

1. They randomly split the data into two data sets: (i) a training data set, and (ii) a testing data set.

2. They use four machine learning algorithms in the training data set to generate a model that predicts the outcome of interest (in our case, application rates or attendance) using *only* baseline data for those in the *control* group.

3. They use the same types of algorithms in the training data set to generate a model that predicts the outcome of interest using baseline data for those in the *treatment* group.

4. They assess the accuracy of the machine learning predictions by validating the models' predictions in the "testing" data set.

5. They produce a predicted "Individual Treatment Effect" (ITE) for people in the "testing" data set by subtracting the predicted outcome from the treatment model with the predicted outcome from the control model.

6. They sort individuals in the testing set by their predicted ITE and then split them into 5 groups. The "top" group has the highest ITEs and the bottom group has the lowest ITEs.

7. They run a regression in the training data set of the actual outcome on a dummy for the treatment, interacted with the five ITE groups, while controlling for the main effect of the ITE group and the values of the two models generated in steps 2 & 3. The regression is weighted based on a propensity score that balances treatment and control groups in each ITE group. This produces what they call "Sorted Group Average Treatment Effects" (GATES).

8. They run steps 1-7 again 100 times, each with a new random split of the original data set into the training and testing sets. This gives ITE and GATES estimates for half of the sample 100 times, or for each observation 50 times in expectation.

9. They take the median value of the coefficients from the GATES regressions and the median value of the standard error.

10. To assess significance of the median GATES estimates at the α level, they choose a conservative critical value associated with $\frac{\alpha}{2}$ instead of the usual *alpha*. This is to take into account the uncertainty associated with the random splits in step 1. For example, for the median GATES estimates to be significant at the 10% level, they require a critical value of 1.96, which is usually the critical value for 5% significance.

We deviate from this procedure in two ways to maximize statistical power. First, in step 9, instead of taking the median value of the GATES coefficients from the 100 simulations, we take the median ITE value for each person in the sample. Due to the nature of the random splitting of the data in step 1, each person in the sample will get an ITE score in about half of the simulations. Taking the median value across all simulations allows us to utilize the entire sample in our regressions as opposed to only utilizing the half of the sample that was put in the "testing sample" split. All of the ITEs for each person are still "honest" in the sense that the estimate is based on data from other people in the sample and is never using the individuals' own behavior. Because we do not have additional uncertainty associated with only using half of the sample, we use the conventional critical values of $1 - \alpha$ in our analysis instead of $1 - \frac{\alpha}{2}$.

The second deviation is in how we assess heterogeneity. Their procedure assesses the degree of heterogeneity by regressing the outcome on the treatment indicator interacted with the individual value of the ITEs. They take the p-value on that interaction to be evidence of heterogeneity. This requires a linearity assumption. We go beyond this by taking a less parametric approach, simply comparing the coefficients from the top GATES group to the coefficients in the bottom GATES group and seeing if they are significantly different. If they are, this is evidence that the top and bottom groups are affected significantly differently by the treatment – in other words, that there is significant heterogeneity in treatment effects. This allows us to assess if there are differences at the extremes, instead of assessing whether there are differences all across the distribution of ITE scores. Since there are more likely to be differences at the extremes, this provides additional statistical power.

To minimize the number of tests we run, we only use the machine learning algorithm that is shown to have the most predictive ability in the testing set. Chernozhukov et al. (2020) use all four machine learning methods (elastic net, boosted trees, random forest, and neural network) each time and report results from the top two.