



European Bank
for Reconstruction and Development

Gender discrimination in small business lending. Evidence from a lab-in-the-field experiment in Turkey

J. Michelle Brock and Ralph De Haas

Abstract

We test for the presence of gender discrimination in small business lending through a lab-in-the-field experiment with 334 Turkish loan officers. Each officer reviews multiple loan applications in which we randomize the applicant's gender. While unconditional approval rates are the same for male and female applicants, we detect a more subtle form of discrimination. Loan officers are 30 percent more likely to make loan approval conditional on the presence of a guarantor when we present an application as coming from a female instead of a male entrepreneur. This gender discrimination is concentrated among young, inexperienced, and gender-biased loan officers. Discrimination is also most pronounced for loans that perform well in real life, making it costly to the bank. Experimental variation in the available applicant information does not impact lending decisions, suggesting that the nature of discrimination is implicit rather than statistical.

Keywords: Gender discrimination; guarantors; implicit bias; lab-in-the-field; bank lending

JEL Classification Number: D81; D83; D91; G21; G41; L26

Contact details: Michelle Brock, One Exchange Square, London EC2A 2JN, UK.

Phone: +44 20 7338 7193; email: brockm@ebrd.com

Michelle Brock is a Principal Economist at the European Bank for Reconstruction and Development (EBRD) and a Research Affiliate at the Centre for Economic Policy Research (CEPR). Ralph De Haas is Director of Research at the EBRD and a Research Fellow at the CEPR.

We thank Martin Acht, Francesca Dalla Pozza, Gözde Esen, Melanie Koch, Victoria Robinson and Alex Stepanov for excellent research assistance and Thorsten Beck, Sergei Guriev, and Sandra Orozco-Aleman, seminar participants at the EBRD and the Surrey Business School, and participants at the Southern Economic Association Annual Meetings for useful comments. The authors gratefully acknowledge financial support from the EBRD Shareholder Special Fund. IRB approval was obtained from the Heartland Institutional Review Board.

The working paper series has been produced to stimulate debate on economic transition and development. Views presented are those of the authors and not necessarily of the EBRD.

1. Introduction

Across the world, many more men than women make use of financial services. Low and middle-income countries in particular remain characterized by large and persistent gender gaps in account ownership and access to bank credit (Demirgüç-Kunt et al., 2018). This study focuses on Turkey, a country where 83 per cent of men but only 54 per cent of women have a bank account. Likewise, 43 per cent of male-owned Turkish firms report being credit constrained while this holds for 63 per cent of female-owned firms (EBRD, 2019).

Whether financial gender gaps entail economic inefficiencies depends on whether they mainly reflect differences in the demand for financial services or different supply-side constraints. Gender discrimination is often cited as a key supply-side driver of women's financial exclusion (The Economist, 2013). Discrimination occurs when lenders treat male and female loan applicants differentially even when they are similar in terms of other characteristics. Such discrimination can undermine market competition as aspiring or existing female entrepreneurs face overly tight credit constraints. The resulting underutilized productive capacity can inhibit potential growth.

Providing rigorous evidence on gender discrimination in access to finance is challenging for at least two reasons. First, not just the supply of credit but also the demand for it can differ by gender (Demirgüç-Kunt, Beck and Honohan, 2008). For example, women may self-select into industries that are less capital intensive or that operate at a smaller scale and hence require less external finance. Empirical attempts to detect gender discrimination must control for such demand differences. Second, gender discrimination can be both direct and indirect. Even when loan officers are unlikely to reject female loan applicants straight away, they may apply gender-specific conditions (such as guarantor requirements) that make credit de facto unattainable for many women. Such indirect discrimination is particularly difficult to spot.

In this study, we test for the presence of both direct and indirect gender discrimination in small business lending in Turkey. We implemented a lab-in-the-field experiment during which loan officers had to evaluate multiple loan applications in which the gender of the applicant had been (randomly) manipulated by us.¹ Our experimental setting allowed us to measure various important loan officer characteristics – such as their implicit gender bias, risk aversion, and work experience – that can be expected to influence biased decision making. We exploit this information to analyze heterogeneous effects across different types

¹ Gneezy and Imas (2017) define a lab-in-the-field study as one “conducted in a naturalistic environment targeting the theoretically relevant population but using a standardized, validated lab paradigm”.

of loan officers. Importantly, loan officers assessed applications that our partner bank had received in the recent past so that we know how loans subsequently performed in reality.

Bringing experienced loan officers into a controlled laboratory environment has three advantages. First, it allows us to track lending decisions carefully. We not only observe whether officers grant credit but also the conditions under which they are willing to lend. We focus in particular on guarantor requirements – where a guarantor or co-signer underwrites the loan – since a previous non-experimental literature suggests that guarantees may be an important source of indirect gender discrimination (Alesina, Lotti and Mistrulli, 2013).² Second, we can exogenously match loan officers with applicants, circumventing issues related to endogenous matching that plague observational studies of gender discrimination. Third, by bringing loan officers to the lab we can complement our experiment with psychometric tests to elicit officers’ risk aversion and implicit gender bias. These lender characteristics, which normally remain unobservable to the econometrician, help us to provide new insights into how discrimination varies within the loan officer population and into the nature of discrimination itself.

We conducted our experiment with 334 loan officers of a large Turkish bank. Turkey provides a particularly interesting setting to study gender discrimination in small business lending. It is a large and growing emerging market with a professional and competitive banking system. The country scores well in terms of *de jure* gender equality: there are few legal obstacles that restrict women’s ability to operate as an entrepreneur (Klapper, Munoz and Singh, 2014). Yet, at the same time, the country remains characterized by very conservative gender norms. It ranks 130 out of 149 countries in terms of *de facto* gender equality (WEF, 2018). This stark discrepancy between gender-related laws on the book and actual attitudes within society is characteristic of many other emerging markets as well.

We find that while unconditional approval rates are similar for male and female loan applicants, loan officers discriminate against women in an indirect way. In particular, officers are 30 per cent more likely to require a guarantor when we present an application as coming from a female instead of a male entrepreneur. This discrimination is concentrated among young and inexperienced loan officers and those that display a high implicit gender bias. Discrimination is also concentrated among loans that performed well in real life, making it

² Unlike “passive” collateral, guarantors actively monitor borrowers to ensure repayment (Banerjee, Besley, and Guinnane, 1994) and monitoring is often leveraged by the threat of social sanctions (Bond and Rai, 2008). The use of guarantors is not only widespread in Turkey and other emerging markets but is also common in many advanced European economies and in the United States.

potentially costly to the bank. Lastly, loan officer decisions are unaffected when we manipulate the applicant information that they have access to. This suggests that the nature of discrimination is implicit rather than statistical.

Our findings allow us to advance the literature on gender discrimination in credit markets in four important ways. First, earlier efforts to detect such discrimination relied on observational data. These papers use multivariate regression analysis in which the dependent variable is a lending outcome and the main independent variable is the gender of the applicant or borrower. Identification relies crucially on controlling for all other applicant characteristics that impact loan outcomes and may correlate with gender. Using this approach, a few papers provide suggestive evidence of discriminatory rejections and loan terms. Bellucci, Borisov and Zazzaro (2010) use data from an Italian bank and show that female entrepreneurs face tighter credit availability and collateral requirements but not higher interest rates. Alesina, Lotti and Mistrulli (2013) access the Italian credit registry and find that female-owned firms *do* pay higher rates (even within the same bank). Women also need to post a guarantee more often even though their firms are not riskier. For obvious reasons, omitted variables bias is a notable drawback of the non-experimental approach. Our lab-in-the-field methodology avoids this problem as we can vary applicant gender while keeping all other characteristics equal.³

A second Achilles heel of studies using administrative data is that they often conflate demand and supply. A lower use of credit by female enterprises can simply reflect lower demand. Some women may be more risk averse, more easily discouraged from applying for credit (Ongena and Popov, 2016), or self-select into industries that need less external finance. Our experimental approach allows us to isolate the supply side of the credit market.

Third, the assignment of applications to loan officers is typically not random in administrative data but reflects observable and unobservable loan officer characteristics. Female applicants may be allocated to experienced or less gender-biased officers, pushing estimates of gender bias downwards. A few studies have addressed this through using matched loan officer-borrower data.⁴ Fisman, Paravisini and Vig (2017) exploit the rotation of loan officers at an Indian bank to show that loan officers increase credit access and reduce collateral requirements for clients of the same religion or caste. In their setting cultural

³ Studies using administrative data from the U.S. generally find no evidence of gender discrimination in small business lending (Blanchflower, Levine and Zimmerman, 2003; Blanchard, Zhao and Yinger, 2008; and Asiedu, Freemant and Nti-Addae, 2012) although they do detect discrimination against African-Americans and Hispanics (Munnell et al., 1996 and Blanchflower et al., 2003).

⁴ See Alesina, Lotti and Mistrulli (2013) and Beck, Behr and Guettler (2013).

proximity helps to reduce information asymmetries rather than instigate favoritism. We go a step further by randomly assigning applications to loan officers so that there is by construction no bias due to endogenous matching.

Fourth, observational data usually lack loan officer characteristics. This limits what we can learn about heterogeneity in gender discrimination. Interesting exceptions are Beck, Behr and Guettler (2013) and Beck, Behr and Madestam (2018) both of which use data from an Albanian microfinance institution. The first paper shows that female loan officers make lending decisions that result in fewer arrears. This advantage only emerges with experience when women lend to male borrowers. The second paper finds, in line with an own-gender bias among officers, that new borrowers matched with opposite-sex loan officers pay higher interest rates. They are also less likely to demand a second loan. The authors of the first paper conclude that “*not only the institutional and governance structure of financial institutions matters, but also the gender of the people operating in a given bank structure*” (p. 5). Yet they acknowledge that a performance gap between male and female loan officers may also reflect unobserved differences. By bringing officers to the lab, we can measure important personality traits that usually remain unobservable, such as their risk aversion and implicit gender bias. This allows for a richer discussion of loan officer heterogeneity and a nuanced view on the role of their gender in particular.

Our results also contribute to the economics literature on discrimination. This literature distinguishes between, on the one hand, taste-based (Becker, 1957) and implicit discrimination (Bertrand, Chugh and Mullainathan, 2005) and, on the other hand, statistical discrimination (Phelps, 1972; Arrow, 1973).⁵ Taste-based discrimination takes place when decision makers (say, loan officers) are prejudiced against a social group (say, women) and consciously avoid interacting with them. Implicit discrimination is similar but occurs without the decision maker being aware of it and is more likely when decisions are made under conditions of ambiguity, time pressure, or distraction (Bertrand, Chugh and Mullainathan, 2005). In contrast, statistical discrimination refers to decision makers relying on a group attribute (such as gender) as a signal of unobserved individual characteristics. Such stereotyping takes place in the absence of reliable individual-level information.⁶ Statistical

⁵ Most of this literature focuses on racial discrimination (for example Neal and Johnson, 1996 and Bertrand and Mullainathan, 2004) and gender discrimination (for example Altonji and Blank, 1999 and Bayard et al., 2013) in the labor market. See Neumark (2018) for a survey of the experimental literature on labor market discrimination.

⁶ While taste-based and implicit discrimination are inefficient, statistical discrimination is theoretically efficient and rational.

discrimination may be especially strong when decision makers evaluate “out-group” people (say, male loan officers screening female loan applications) because it can be difficult to interpret signals from such out-groups (Cornell and Welch, 1996). Distinguishing between these types of discrimination is challenging as they can occur simultaneously and exacerbate each other. For instance, if decision makers collect less information about individuals in certain groups due to taste-based or implicit discrimination, this may over time result in statistical discrimination (Bertrand and Duflo, 2017).

Our contribution to this literature is to use a lab-in-the-field experiment to not only measure discriminatory behavior, but also to uncover the nature of any detected bias. We do this by experimentally varying the information loan officers have access to; assessing the real-life quality of lending decisions; and eliciting implicit gender bias through an implicit association test (IAT). We find that variation in the available information does not affect lending behavior; that loan officer heterogeneity in implicit gender bias correlates with discriminatory behavior; and that discriminatory lending decisions do not boost loan quality. We therefore conclude that the gender discrimination that we observe is likely to be of an implicit rather than statistical nature. We also find that loan officer experience reduces discriminatory lending, possibly because experienced officers better contain any implicit gender biases they may harbor.

Lastly, our paper also provides a methodological contribution since we integrate several features of earlier experimental work within one setting. Our general approach is a lab-in-the-field experiment in which we observe actual loan officers (rather than, say, students) working on a task that closely resembles their everyday work.⁷ In the first part, we randomly manipulate the gender of loan applications as is typical of this approach.⁸ Yet, unlike correspondence studies that use fictitious applicants, we use real-world loan applications. This ensures that the applications are realistic and that we can observe what happened to them in real life (Cole, Kanz and Klapper (2015) follow a similar approach). Moreover, we properly incentivize all decision making in our experiment to mimic the costs that may be associated with certain forms of discrimination. This contrasts with most existing experimental work on discrimination, which typically does not impose costs on participants for making inefficient discriminatory choices in hypothetical scenarios (Neumark, 2018).

⁷ Our experimental design could also be described as a two-pronged lab-based correspondence study.

⁸ In their seminal correspondence study, Bertrand and Mullainathan (2004) use fictitious resumés with randomly assigned African-American or White-sounding names that are sent in response to a job ad. They find that White-sounding names receive 50 percent more call-backs for interviews.

In the second part of the experiment, we randomize the type of information that loan officers can access. When a correspondence study is rich in applicant characteristics, statistical discrimination is less likely (Neumark, 2018). Vice versa, withholding some information may increase gender differences in loan conditions. We would interpret such a finding as evidence for statistical discrimination in which loan officers rely more on applicants' group membership (gender) whenever less individual information is available.⁹

Lastly, we elicit loan officers' risk aversion and implicit gender bias. Combining correspondence studies with IATs is still rare. Rooth (2010) implements an IAT among recruiters in a correspondent study on discrimination against Arab-Muslim men in Sweden. More closely related to this paper, Reuben, Sapienza and Zingales (2014) conduct a lab experiment in which "employers" hire "workers" for a mathematical task. They find that employers are more likely to hire males when they have no information about the ability of hires. This pattern is reduced, but not eliminated, when information on prior mathematical performance is made available. The researchers also administer an IAT based on stereotypical associations between gender and words related to mathematics/science versus liberal arts. They find more bias, and less updating based on performance data, for those who are more gender biased according to the IAT. Implicit biases may thus interact with information availability and statistical discrimination.¹⁰

The paper is structured as follows. Section 2 describes our setting and experimental design. Section 3 then summarizes the data generated by the experiment and outlines our estimation strategy. Section 4 presents the results and Section 5 concludes.

2. Experimental context and design

2.1. Context: Participant population and loan application process

We conducted our experiment in cooperation with a large private commercial bank in Turkey. We held 22 experimental sessions with a total of 334 bank employees across eight cities over a two-month period.¹¹ The bank operates a regional office in each of these cities

⁹ Kaas and Manger (2012) use a correspondence study in Germany to study the effect of Turkish-sounding names on call back rates for job interviews. Discrimination was eliminated when a reference letter, containing indirect information about productivity (such as applicants' conscientiousness and agreeableness) was added. The authors interpret this evidence as being consistent with statistical discrimination.

¹⁰ A meta-analysis of 184 studies found an average correlation of 0.24 between the IAT score and outcome measures such as judgments, choices, and physiological responses (Greenwald et al., 2009).

¹¹ These cities were Adana, Ankara, Antalya, Bursa, Gaziantep, Istanbul, Izmir, and Trabzon. We conducted a pilot with 32 employees in Istanbul but do not use these data.

and participants were randomly selected from all bank employees involved in SME lending (two-thirds of the bank's loan portfolio consists of SME credit). Figure 1 shows the location of the regional offices and the number and gender of the participating bank employees.¹²

[Insert Figure 1 here]

Our participant sample consists of two types of bank employees: loan officers (192) and supervisors (142).¹³ The loan application process at our partner bank is highly standardized. Loan officers establish contact with a new (potential) borrower and conduct the initial screening. They interview the client, fill out a site-visit form and collect documentation on business performance (income statements and balance sheets). They also check the availability of any collateral and guarantors and request a credit score from the Turkish credit registry (KKB). Loan officers enter this objective screening information into an electronic application form. At this stage they can also voluntarily add more subjective notes to the form, such as about the client's perceived trustworthiness¹⁴, experience or social standing.

If the loan officer deems a client creditworthy in principle, they pass on the electronic application form to their supervisor (typically the branch manager) with a proposal about the maximum credit limit. Crucially, at this point they also include their view as to which collateral and guarantors (if any) are required. That is, loan officers can recommend that the loan application is approved unconditionally or conditional on the presence of collateral, in particular a guarantor. The supervisor reviews the loan application and can reject or provisionally approve it. In the latter case, they need to send it to the bank's headquarters for formal sign off.¹⁵ Both loan officers and their direct supervisors are thus based in the same regional office and actively engage in the screening of potential borrowers.¹⁶

¹² We stratified by gender, so that the participants' gender composition does not exactly reflect that of the local universe of bank employees.

¹³ Our paper therefore also relates to recent work on how the organizational and incentive structure of banks affects loan officers' decision making (for example Hertzberg, Liberti and Paravisini, 2010; Berg, 2015; Cole, Kanz and Klapper, 2015; Qian, Strahan and Yang, 2015; and Agarwal and Ben-David, 2018).

¹⁴ See Duarte, Siegel and Young (2012) on the importance of trust in the context of on-line peer-to-peer lending. Borrowers who appear trustworthy are more likely to obtain a loan and pay a lower interest rate (which is in line with the actual lower default probability of such borrowers).

¹⁵ Branches can approve loans below a certain size threshold but in practice only 10 (0.5) per cent of micro (SME) loans are formally signed off in a branch. Micro clients are those with an annual turnover below TRY 2.5 million (\approx US\$ 700k) and a credit limit below TRY 750k (\approx US\$ 210k). The application process is fast, with applications typically approved within 1.5 days.

¹⁶ Henceforth we refer to the total experimental population as either "participants" or simply "loan officers".

We study whether loan officers treat female and male loan applicants differently, all else being equal. An important contextual question is whether there are average differences between the creditworthiness and performance of male and female small businesses in Turkey. We provide three pieces of evidence that suggest this is not the case. First, the rate of non-performing loans (NPLs) among female and male-owned SMEs is 2.9 and 2.7 per cent, respectively, in the loan portfolio of the bank we work with. This difference is not statistically significant.

Second, we randomly sample 250 loan applications (stratified by gender) out of all applications the bank received from small firms in recent years. We then compare the credit registry score of male and female applicants (a higher score implies less risk). While the average score is 1,035 for men and 1,023 for women, this difference is not statistically significant ($p=0.80$). Appendix Table A2 presents linear probability regressions for the 243 files for which these credit scores were available (the dependent variable). The first column confirms there is no significant difference between female and male applicants. This holds when we include sector fixed effects (column 2) or sector and region fixed effects (column 3) and when we control for firm size (column 4) and the loan amount requested (column 5).

Third, we follow Ongena and Popov (2016) and run regressions using the Turkish Business Environment and Enterprise Performance Survey (Appendix Table A3). The dependent variable is firm-level sales growth over the past three years. The main explanatory variable, *Constrained female vs unconstrained male*, is a dummy that is one if the firm owner is female *and* was rejected when she applied for a bank loan or was discouraged from applying for one. This dummy is zero if the owner is a man whose loan application was approved. The results indicate that when conditioning on firm covariates as well as industry and year fixed effects, there is no significant difference in sales growth between *constrained* female-owned firms and *unconstrained* male-owned firms. In short, male and female-owned Turkish small businesses display very similar profiles in terms of risk and performance.

2.2. *Experimental design*

For the main task in our experiment participants evaluated two rounds of four loan applications (eight in total).¹⁷ Applications were randomly presented as coming from a female or a male entrepreneur. Participants had to decide whether to approve or reject each application and, in case of initial approval, whether to request a guarantor or not. For each

¹⁷ Loan officers made decisions on loan applications worth US\$ 81.1 million in total.

loan application, participants also had to provide a subjective risk rating between 0 and 100. We did not constrain the time participants had to evaluate each application.

The task closely mimicked the choices the participants make in their daily work. To be consistent with day-to-day lending decisions, all loan application material was presented to the participants electronically and in the format of the standard application forms that bank employees normally process on their own computer. The forms (henceforth called “loan applications”) contained all the information available at the time the application was processed.¹⁸

We use a total of 100 unique loan applications in the experiment, which we selected from an initial sample of 250 applications. These 250 applications were a stratified random sample of all applications by existing SMEs (that is, no start-ups) that the bank had received in the three to six years before the experiment.¹⁹ Using this earlier period allows us to track what happened to each of these applications in real life. The strata were region, gender, firm size, and whether the application was accepted in real life. By only using applications from first-time loan applicants, which had never before borrowed from our partner bank, we eliminate the potential influence of soft information generated over time.

All applications used in the experiment were gender neutral except for the randomly assigned applicant name. To achieve this, we only used applications from Turkish industries in which both female and male ownership is relatively common. For instance, we excluded applications from construction companies (a male dominated sector) and beauty parlors (a female dominated sector). Some files were eliminated because they contained explicit references to the applicant’s gender (other than their name) or because they were incomplete. In this way we identified the 100 gender neutral files used in the experiment.

All 100 files occur in the data multiple times, sometimes as coming from a male applicant and sometimes from a female applicant. This allows us to obtain a within-application estimate of gender discrimination. By asking participants to review both male and female

¹⁸ These forms are at the heart of the decision making about whether the bank is willing to lend, what the maximum exposure is, and whether a guarantor is required. Only after this stage, do the loan officer and client negotiate about specific product types, such as credit lines and fixed-term loans. The maturity and pricing of individual products is also determined at this later stage. This implies that during the experiment we collected data on willingness to lend, maximum amount granted and the need for a guarantor, but not on the interest rate and maturity of specific credit products.

¹⁹ When participants evaluated the files, they did not see the real application date but instead a date in the year of the experiment. We presented files as contemporary to avoid recall bias – loan officers did not have to think back about the economic situation in the past. This adjustment did, of course, lead to a slight disconnect between loan performance in real life and the loan application evaluated during the experimental session.

applications, we preserve external validity as no one at the bank sees only male or female clients. We also held constant the ratio of performing, non-performing and rejected files each participant saw, at 2-1-1. This 2-1-1 ratio does not reflect the bank's actual application flow, but we used this ratio so that participants saw at least one file of each type. Each participant also evaluated both high- and low-quality applications from men and women.²⁰ Due to time constraints participants could not evaluate more than four files and we wanted to ensure the data could be analyzed by application quality.

We determined the gender of each file by assigning new names, randomizing between male ones (Ahmet, Ali, Mehmet, Mustafa) and female ones (Ayse, Emine, Fatma, Zeynep). These names were chosen because they are common across Turkey and are well represented among working-age adults across regions.²¹ Names were randomized such that each participant saw one performing loan and one "bad" loan application (either a non-performing loan or a declined application) from each gender. Each loan application was on average evaluated by 13.4 different participants per round, half of the time as a female and half of the times as a male file.

After a first round of four loan applications, lending staff were presented with a second round of four applications. We again randomized the gender of each application. Moreover, we now also randomized participants into three groups: one control group evaluated applications with all information available (as in the first round), a first treatment group evaluated files from which we had deleted the credit score from Turkey's credit bureau, and a second treatment group evaluated files where a section with subjectively provided information had been removed.²² This section contains voluntary comments by loan officers about the applicant (such as about how industrious they are or whether they have a good business network). Bank staff provide this information in the application file to strengthen the rationale for lending to the applicant in question. It targets the decision makers higher up in the lending hierarchy. If either the objective credit score and/or the more subjective comments section contributes to staff's ability to make fair and objective lending decisions, omitting it may increase statistical discrimination, especially among loan officers with an

²⁰ That is, analogous to Bertrand and Mullainathan's (2004) correspondent study on racial discrimination, we crossed applicant gender with application quality.

²¹ We checked which names had the highest frequencies in the relevant cohorts and across regions using information from the Turkish General Directorate of Population and Citizenship Affairs (<https://www.nvi.gov.tr/isim-istatistikleri>) and an additional online data source (<https://www.isimarsivi.com/>).

²² We did not present a blank version of these sections but instead removed the entire section. We did this to avoid the possibility that an empty section unintentionally acts as a quality signal. This, in turn, would make it more difficult to interpret the treatment effects properly.

implicit gender bias (Reuben, Sapienza and Zingales, 2014). On the other hand, if the information itself is perceived with bias, omitting it may potentially reduce discrimination. In the first (second) case, we should see that bias is higher (lower) in the treatment groups than in the control group.²³

For the second round, we opted for a within-file (in terms of gender randomization) and between-participant (in terms of the information treatment) experimental design for two reasons. First, we wanted to avoid non-linear or heterogeneous order effects. Non-linear order effects are difficult to control for, while controlling for heterogeneous order effects would require a larger participant pool than we had. Second, asking all participants to do all treatments would have required each participant to complete 12 reviews, and there was not sufficient time for such an extensive exercise.

We incentivized all loan decisions in line with the structure of common bank incentive schemes. Participants earned ten points (equivalent to ten Turkish lira) for each completed review (quantity) and an additional five points when they correctly approved a loan that performed well in real life (quality).²⁴ In contrast, five points were deducted when they incorrectly accepted a loan that was defaulted on in real life. When participants approved a file that had been declined in real life, we gave them a 50/50 chance that the file was counted as performing, thus yielding the extra five points. We did not penalize incorrect rejections in order to mimic as closely as possible the actual incentive scheme at the bank, and the bank cannot realistically know when a rejection is incorrect.

The total number of points per participant was aggregated at the end of the experiment. Participants were then ranked according to their total score and split into four quartiles. Participants did not learn how many points they earned but were told their quartile. This corresponds with our instructions at the beginning of the session, when we informed them that their ranking would determine the final rewards. Those in the highest quartile could spend their points on a selection of higher valued prizes while those in the lower quartiles had to select gifts with lower values. All participants had chosen their preferred prizes from each category prior to the experiment. This ensured they understood how the incentives worked

²³ Heilman (1984) documents inconsistent findings in past experimental studies as to whether giving additional information about individual applicants reduces adverse outcomes for women.

²⁴ This incentive scheme resembles the remuneration system that the bank uses in reality and is also similar to the low-powered (baseline) scheme of Cole, Kanz and Klapper (2015).

and what the benefit would be of getting into the top quartiles.²⁵ The incentive scheme was thus ultimately both material and competitive.

2.3. Eliciting personality traits

After both rounds of application decisions, we implemented two experimental modules to measure participants' risk preferences and implicit gender bias. We first follow Binswanger (1982) and Eckel and Grossman (2008) and elicit risk preferences by presenting participants with six stylized risk scenarios from which they had to choose one. Each scenario was depicted as a circle split in two: left and right. Each half contained a possible point outcome and the even split represented that in each scenario there was a 50 per cent chance of getting the left-hand or right-hand outcome. The outcome pairs were 28-28; 20-44; 24-36; 16-52; 12-60; and 2-70. The first scenario has no risk as the numbers are equal while the last scenario entails most risk. The task was incentivized: an on-site computer drew random draws to determine whether participants would get the low or high number from the circle they selected. We added the resulting points to participants' point total at the end of the session.

Second, participants took an implicit association test (IAT).²⁶ The IAT began with a brief survey to obtain key demographic and personal characteristics such as professional experience, education, age, and gender. Participants were then introduced to the association task, which is the core of the IAT, and had two practice rounds. The association task consists of sorting, as quickly as possible, words that appeared sequentially on their tablet. For example, in one task "female" words had to be allocated to the category "family" and "male" words to the category "career" (the stereotypical task).²⁷ In another task, "male" words had to be allocated to the category "family" and "female" words to "career" (the non-stereotypical task). The time a respondent took to allocate each word was recorded in milliseconds. The presumption is that respondents with a stronger association between two concepts find the

²⁵ Specimens of the possible prizes that participants could buy with the points they earned during the experiment were on display during the sessions. The actual prizes were sent to the participants a few weeks later.

²⁶ See <https://implicit.harvard.edu/implicit/takeatest.html>, Carney et al. (2007) and Lane et al. (2007). IATs are by now common in both psychology (Greenwald, McGhee and Schwartz, 1998) and economics (Bertrand, Chugh and Mullainathan, 2005; Beaman et al., 2009; Carlana, 2019). While attitude IATs are used to detect implicit (unconscious) negative attitudes towards social groups, stereotype IATs measure implicit associations between social groups and specific traits (Bertrand and Duflo, 2017). Our IAT falls in the latter category.

²⁷ The IAT and all other documentation was provided in Turkish. The family-related words were the Turkish translations for "garden", "kitchen", "marriage", "laundry", "home", "children" and "relatives". The career words were "office", "manager", "income", "job", "profession", "employees" and "business". To designate "male" we used "man", "he", "men", "him", "boy", "his", "gent" and for "female" we used "woman", "she", "women", "her", "girl", "hers", and "lady".

sorting easier and complete it faster in one task compared to the other. We defined a participant's implicit bias as the normalized difference in mean response times between the non-stereotypical and stereotypical task. Higher values indicate stronger bias.

3. Data and estimation strategy

3.1. Data

Table 1 summarizes our experimental data (Appendix Table A1 contains all variable definitions). Panel A describes the main characteristics of the 334 participants. Almost half of them are female and their average age is 37 years, ranging between 26 and 53. Forty-three per cent of the participants are supervisors, the others are loan officers. There is wide variation in participants' lending experience. While the average participant has worked at the bank for five years, this varies between zero and 19 years.²⁸ There is thus substantial heterogeneity in the lending experience that loan officers built up over the course of their career. Learning through experience can mitigate or even eliminate statistical discrimination over time (Altonji and Pierret, 2001) especially for risk-averse decision makers (Aigner and Cain, 1977).²⁹

[Insert Table 1 around here]

Unlike in observational studies, our lab-in-the-field experiment allows us to measure participant characteristics that are otherwise difficult to observe. As described in Section 2.3, we use lottery questions to elicit risk aversion and an IAT to gauge participants' implicit bias against women in business. Table 1 reveals substantial variation in these measures as well. The categorical variable *Participant risk aversion* ranges between 1 (risk loving) and 6 (most risk averse). The average participant scores 4.1. A large literature has documented that, on average, women tend to be more risk averse than men (for example Eckel and Grossman, 2008). More specifically, evidence from the financial-services industry indicates that female decision makers take less risk on average (Sunden and Surette, 1998; Agnew, Balduzzi and Sunden, 2003). The correlation matrix in Appendix Table A4 shows that this is the case in our setting as well. The average risk aversion score is 4.32 for females and 3.92 for males and this difference is statistically different at the 10 per cent level.

²⁸ We also asked participants whether they had previous lending experience in a similar role at another bank. All results that follow are robust to using this broader experience definition.

²⁹ Moreover, earlier (non-incentivized) experimental evidence indicates that more experienced loan officers acquire more financial information when lending to fictitious small firms (Andersson, 2004).

The IAT score is normalized between -1 and 1 with zero indicating no implicit gender bias. While the scores range widely (between -0.93 and 1) a large majority of lending staff (87 per cent) has a positive IAT score, indicating that they subconsciously associate business more with men than with women. This tendency is stronger among women than among men. Figure 2 shows that the IAT distribution for female participants lies to the right of that for men. A two-sample Kolmogorov-Smirnov test confirms that the distributions are significantly different (see also Appendix Table A4). The average IAT score is 0.39 for women and 0.28 for men and this difference is also statistically significant at the 5 per cent level. The IAT scores of female participants are thus on average almost 40 per cent higher than those of men.

[Insert Figure 2 around here]

Panel B of Table 1 summarizes the real-life characteristics of the 100 loan files that we use in the experiment. By design, half of these loan files refer to loans that in real life were paid back on time (performing loans), a quarter refers to loans that in real-life were defaulted upon (non-performing loans), and another quarter consists of loan applications that were rejected in real life (declined applications). Panel C summarizes the experimental outcomes at the participant-file decision level. We show separate statistics for round 1 (no information treatment) and round 2 (information treatments). In both rounds, almost forty per cent of the loan applications were rejected whereas, conditional on acceptance, a guarantor was requested in 27 per cent of the cases. Panel D shows that we withheld either subjective or objective applicant information in a third of the decisions in this second round.

For each credit application, the participant was asked to estimate, on a 0-100 scale, the probability that the borrower would repay. This helps us to verify that the experimental task was meaningful in the sense that loan officers could infer credit risk based on the information in the loan file. Figure 3 provides a scatterplot of the 100 files used in the experiment, based on data from the first round only. The horizontal axis indicates the average subjective repayment probability (each file was evaluated by 13.4 participants on average in round 1) while the vertical axis shows the share of participants that rejected the application. Figure 3 reveals a tight negative correlation between expected repayment probability and the likelihood of loan rejection. This suggests that our incentive scheme worked and that participants thought the task realistic and paid attention to the information provided.

Equally important is to check whether the decision making in our lab-in-the-field correlates with what happened to the loan applications in real life. We find that this is the case. Overall, 72 per cent of all applications that related to loans that performed well in real life were approved during the classroom sessions. This percentage is significantly lower for applications related to non-performing loans (53 per cent) and for applications that were rejected in real life (47 per cent). As a result, files that in real life were non-performing (red dots) or declined (orange) are concentrated in the upper left corner of Figure 3 while performing loans (green) are concentrated in the lower right-hand corner. This indicates that across the board participants correctly identified loans that performed well or badly in real life and made lending decisions in line with these subjective perceptions of loan quality.

[Insert Figure 3 around here]

Appendix Table A4 provides a correlation matrix of the participant characteristics and the rejection dummy. We have already discussed that female participants are on average more risk averse and more gender biased. There are two other interesting correlations in this table. First, participants higher up in the lending hierarchy (supervisors) are more often female, older and more experienced, as well as more gender biased. Second, older participants (unsurprisingly) tend to be more experienced but also more gender biased, suggesting that cohort effects are important. Lastly, it is reassuring that whether a file was presented as male or female (*Female applicant*) is uncorrelated with any of the participant characteristics. This reflects that the randomization process was successful.

Appendix Table A5 assesses the correlates of implicit gender bias in a multivariate setting. When we “horse race” the participant characteristics in this way, participants’ own gender is the main variable that explains implicit gender bias. Even when controlling for a participant’s experience, age, hierarchical position, and risk aversion, we continue to find that female bank employees are on average 0.124 points (on the [-1,1] scale) more biased against female entrepreneurs as compared with male bank employees.

3.2. Estimation strategy

To test for the presence of gender discrimination in the lending behavior of bank employees, we run linear probability regressions at the decision level:

$$y_{il} = \alpha + \beta \cdot G_{il} + \sum_{k=1}^K \gamma_k \cdot X_i + \varphi_l + \varphi_c + \varepsilon_{il} \quad (1)$$

Where y_{il} is a lending outcome of interest when participant i evaluates loan application l ; α is a constant; G_{il} is the randomly assigned applicant gender to loan application l as seen by participant i ; X_i is a vector of K participant characteristics (gender, experience, age, a *Supervisor* dummy, risk aversion, and IAT score); φ_l are loan application (file) fixed effects; φ_c is a set of fixed effects for the eight cities where the experiment took place; and ε_{il} is a stochastic error term.

Standard errors are heteroskedasticity robust. Due to the experimental design, applicant gender is the only loan-application characteristic that (randomly) varies across decisions about the same application. The loan application (file) fixed effects thus absorb all other observed and unobserved file characteristics. Because we randomize applicant gender at the file level, within-cluster correlation in the treatment is zero and there no need to cluster standard errors. The standard robust variance estimator yields correct inferences (Abadie, Athey, Imbens and Wooldridge, 2017).

4. Results

4.1. Baseline results

Table 2 presents linear probability regressions based on Equation 1 where the dependent variable, *Rejection dummy*, is “1” if a loan application was rejected by a participant and ‘0’ if it was approved. The independent variable of interest, *Female applicant*, is a dummy whether the application was presented as coming from a female (“1”) or male (“0”) entrepreneur. All specifications include application (file) fixed effects, city fixed effects, and the following participant covariates: *Participant is female*; *Participant experience*; *Participant age*; and whether they are a supervisor or loan officer (*Participant is supervisor*). In columns 2 and 3 we also control for *Participant risk aversion* and implicit gender bias (*Participant IAT score*), respectively. Column 4 includes both. All data are from the first experimental round only.³⁰

Table 2 shows that we cannot reject the null hypothesis of no significant treatment effect of *Female applicant* on loan rejection. Across the four specifications, the coefficient for *Female applicant* is close to zero and, if anything, negative. Since we include file fixed

³⁰ All results also hold when we combine the observations from the first round with those from the control group of the second round (in which we did not delete any information).

effects, our results show that *one and the same* loan application does not have a higher chance of being rejected when we present it with a female rather than a male applicant name. In short, we find no evidence of direct gender discrimination.

[Insert Table 2 here]

Turning to the participant covariates, we find that older participants are slightly less likely to reject loan applications, even when controlling for their experience.³¹ This holds at the 10 per cent level of statistical significance. Moreover, participants who in real life are supervisors are 10 percentage points less likely to accept a loan application as compared with loan officers. This is a large difference as the unconditional acceptance rate in our experiment is 61 per cent. This finding is significant at the 1 per cent statistical level and likely reflects that the main role of supervisors is to validate (or overrule) the initial lending decisions made by loan officers.³² Lastly, we do not find that participant risk aversion or their implicit gender bias have an independent effect on the rejection probability. We include columns 2 and 3 to illustrate that this lack of significance is not due to multicollinearity. Thus, while implicit gender bias is widespread (and varied) among our loan officer population, according to the IAT scores, this bias does not translate into explicit discrimination in lending decisions.

In Table 3 we assess whether the null result is driven by loan officer heterogeneity. We do not find evidence of heterogeneous effects. We cut the data in six ways – by participant gender, above/below median experience, above/below median age; supervisors versus loan officers; above/below median risk aversion and above/below median implicit gender bias – and provide sample-split regressions. We only present the coefficient for *Female applicant* but all regressions include the same covariates and fixed effects as in Table 2. There is no evidence of direct gender discrimination in any of these sample splits. None of the *p*-values for formal Wald tests for the equality of coefficients in column pairs indicate differences at conventional levels of statistical significance.³³

³¹ Both variables are positively correlated ($p=0.50$). See Appendix Table A4.

³² Appendix Table A6 shows that this higher conservatism among supervisors is independent of the (randomized) gender of the loan applicant: the interaction between *Female applicant* and *Participant is supervisor* is never statistically significant. The same holds when we run separate regressions for split samples of loan officers and supervisors.

³³ When we partition continuous variables, the below-median sample contains values strictly below the median while the above-median sample contains values at the median and above. All results remain unchanged when we instead allocate at-the-median observations to the below-median group or when we partition at other points.

[Insert Table 3 here]

4.2. Applicant gender and guarantor requirements

Our results so far indicate that loan officers treat loan applications in the same way regardless of whether we present them as coming from male or female entrepreneurs. Moreover, not only do we detect no gender discrimination in unconditional approval decisions among the loan officer population as a whole, the same holds for various sub-populations. We next assess whether there exist more indirect forms of discrimination against female applicants. In particular, an earlier (non-experimental) literature suggests that the approval of female loan applications is more often conditional on the presence of a guarantor. Such indirect discrimination can negatively impact female entrepreneurs that do not have access to such a guarantor. It can also be disadvantageous to the bank if such entrepreneurs are, in fact, good credit risks. Our experimental set up allows us to investigate both these dimensions.

We start in Table 4 by analyzing whether loan officers are more likely to ask for a guarantor when the application comes from a female instead of a male entrepreneur (all else being equal). The structure of the table mirrors that of Table 2. The sample is smaller as the decision to require a guarantor is conditional on initial loan approval. We find strong evidence of indirect gender discrimination: loan officers are seven percentage points more likely to make final loan approval conditional on the presence of a guarantor when the application is shown to them as coming from a female instead of a male entrepreneur. This is a substantial effect as only 27 per cent of all pre-approved loans are required to have a guarantor. Unlike the case for loan rejection, none of the participant characteristics impact the probability that a guarantor is required.

[Insert Table 4 here]

In Table 5, we present the heterogeneity analysis for unequal guarantor requirements, similar to Table 3. We find a consistent and intuitive pattern of statistically significant heterogeneous treatment effects. When we present the application as coming from a woman instead of a man, loan officers are more likely to ask for a guarantor when they are less experienced (columns 3-4); younger (columns 5-6); and/or display more implicit gender bias during our IAT (columns 11-12). For instance, loan officers with a below-median level of lending experience are 11 percentage points more likely to ask women (as compared with men) for a guarantor than their more experienced colleagues. Likewise, loan officers with

above-median levels of implicit gender bias are 12 percentage points more likely to demand a guarantor when we present a file as coming from a female rather than a male entrepreneur. This more subtle form of gender discrimination is also concentrated among loan officers at the lower hierarchical level, that is, the lending staff who in real life make the initial screening decisions (columns 7-8). This indicates that businesswomen with good loan propositions, but without a guarantor, may be rejected at the initial screening stage.³⁴

Importantly, columns 1 and 2 of Tables 3 and 5 show that there is no difference between male and female participants in how they treat female applicants.³⁵ There is also no statistically significant difference between participants with higher versus lower levels of risk aversion, a characteristic strongly correlated with participant gender (columns 9 and 10). This implies that earlier evidence based on observational studies (such as Beck et al., 2013), and which suggested that male and female loan officers treat female loan applicants differently, may at least partly reflect deeper personal characteristics that are usually left unobservable (such as implicit gender bias) rather than loan officers' gender *per se*. This more nuanced view has important policy implications to which we return in the conclusions.

[Insert Table 5 here]

4.3. Indirect gender discrimination and loan quality

The applications that loan officers reviewed during the experiment were real applications that had been processed by the bank in the recent past. We therefore know what happened to these applications: whether they were rejected or approved and, if approved, whether the loans were repaid or not. We already know from Figure 2 that applications that were rejected or that turned out to be non-performing loans in real life were less likely to be approved during the experiment. In contrast, applications that were approved and subsequently performed well in real life were more likely to be accepted during the classroom sessions.

We now ask whether the higher probability that female loan applicants are required to have a guarantor (relative to male applicants) is driven by loans that performed well in real life or by those that did less well. Figure 4 gives a nonparametric answer to this question. We divide all loan applications into those that were accepted in real life and performed well (left-

³⁴ Note that while supervisors are more conservative than loan officers in terms of approving loans, their conservatism is “gender neutral”. In contrast, when it comes to guarantor requirements, loan officers are not only more conservative than supervisors but there is also a bias against women.

³⁵ This holds independently of whether we control for other participant characteristics or not.

hand side); those that were accepted and became non-performing (middle) and those that were rejected in real life (right-hand side). The bars in the middle and at the right show that for lower-quality applications it does not matter whether we present them as coming from male or female entrepreneurs. The likelihood that a guarantor is requested is not significantly different by gender (if anything, it is slightly higher for male applicants). In sharp contrast, however, the first two bars show a large and statistically significant ($p=0.01$) gender difference. For applications that were accepted and subsequently performed well in real life, an entrepreneur is 9.9 percentage points more likely to be asked for a guarantor when we present the application as female (30.0 per cent) rather than male (20.1 per cent). That is, it is for the best-performing category of loans that loan officers discriminate between male and female applicants in terms of guarantor requirements.

[Insert Figure 4 here]

In Table 6 we perform this analysis parametrically. Column 1 shows that when we control for loan officer covariates as well as file and city fixed effects, women are 12.4 percentage points more likely to be asked for a guarantor in case of high-quality loans (column 1). This difference is absent for loans that were either rejected or non-performing in real life (column 2). The indirect gender discrimination that loan officers display in terms of requesting guarantors is thus not driven by the lower-quality segment of the application pool. Instead, discrimination occurs in the case of relatively good loans that were paid back in real life.

As in Table 5, we can assess which participant types are responsible for this gender discrimination. Columns 3 to 14 show similar heterogeneity as before. High-quality female loan applications are about 15 percentage points more likely to be asked for a guarantor compared to an identical male loan application if the participant is relatively inexperienced (columns 5-6); relatively young (columns 7-8); a loan officer rather than a supervisor (columns 9-10); relatively risk averse (columns 11-12); and has a relatively high implicit gender bias (columns 13-14).³⁶ The difference in the size of the sub-sample coefficients is large throughout Table 5 but less precisely estimated as a result of the smaller sample (performing loans only).

³⁶ There is, again, no difference by participant gender: the coefficients in columns 3 and 4 are not significantly different.

In summary, the discriminatory behavior of less experienced and more biased loan officers in terms of collateral requirements is not concentrated among low-quality loans but instead affects high-quality applications that represent loans that performed well in real life.

[Insert Table 6 here]

In unreported further analysis, we assess whether the (randomized) gender of the applicant had an impact on the ability of loan officers to correctly identify loans that were non-performing in real life (that is, avoiding type II errors) and/or to correctly identify loans that performed well in real life (avoiding type I errors). Overall, we find that applicant gender does not affect either form of prediction accuracy. This holds for all types of loan officers. The one exception is that loan officers are slightly less successful in identifying performing files when we present these as coming from a female entrepreneur. This effect is driven by relatively inexperienced loan officers. This is in line with the results in Table 6 showing that inexperienced loan officers are more likely to require a guarantor from women in the case of loans that performed well in real life.

4.4. Applicant gender and the availability of applicant information

Table 7 assesses whether restricting the information that is available to loan officers has a disproportionate impact on female loan applications, all else being equal. In columns 1 and 2 (3 and 4), we present linear probability regressions where the dependent variable is our *Rejection dummy* (*Guarantor dummy*). In specifications 1 and 3, we include dummy variables that indicate whether in a particular decision we randomly withheld subjective (*No subj.*) or objective (*No. obj.*) loan application information. The former refers to subjective information voluntarily entered by loan officers at the earliest stage of client contact. The latter is the credit score from the Turkish credit registry. All specifications include our standard participant covariates and fixed effects.

The main take-away from Table 7 is that varying the available objective and subjective applicant information does not affect gender discrimination. The interaction terms between *Female loan applicant* and the information treatments are not statistically significant in columns 2 and 4. We do find, however, that withholding subjective borrower information increases the likelihood of loan rejection by 6 percentage points (column 1). Yet, this effect does not differ between male and female loan applicants (column 2). The subjective

information that loan officers can voluntarily add to an application file thus appears useful in the sense that it increases the willingness to lend of those who subsequently review the file.

The fact that experimental variation in the amount of applicant information does not impact loan officer behavior in a gender-specific way, suggests that discrimination is implicit rather than statistical in nature. This is also corroborated by our earlier finding that gender discrimination in terms of guarantor requirements is concentrated among loan officers with a high implicit gender bias.

[Insert Table 7 here]

5. Conclusions

Across a wide range of emerging market economies, women are less likely to use formal financial services. A policy question of first order importance is whether this gender gap reflects supply constraints in general and lender discrimination in particular. To the extent that this is the case, a lack of finance may be preventing women from exploiting their entrepreneurial potential and holding back economic convergence between poorer and richer parts of the world.

To assess the empirical relevance of gender discrimination in small-business lending, we implemented a controlled, randomized lab-in-the-field experiment with a large sample of Turkish loan officers. The loan applications assessed in the experiment consisted of real, previously processed applications. Accordingly, we were able to pay participating loan officers performance incentives based on their lending decision and the realized outcome of the loans they approved. Because gender was randomly (re-) assigned to each application, the same application was sometimes linked to a male and sometimes to a female entrepreneur. As a result, any differences we find can be fully explained by these gender differences.

While we do not find evidence of differences in unconditional approval rates for men and women, there exists a more subtle form of discrimination. All else being equal, the approval of female applications is up to 30 per cent more likely to be made conditional on the presence of a guarantor. This gender discrimination varies with loan officer traits such as their age, experience, and implicit gender bias. Discrimination is also exclusively concentrated among loans that in real life performed well, making it costly to the bank.

A number of loan officer traits – their experience, risk aversion and implicit gender bias, – turn out to be strongly correlated with indirect gender discrimination. These deeper (and usually unobservable) traits appear to be much more important than loan officers' gender, at least in our context. For instance, while we find that female loan officers are significantly

more gender biased on average, it is this variation in gender bias that drives indirect discrimination rather than loan officers' gender per se. These results warn against drawing overly strong conclusions about the relevance of loan officer gender for lending outcomes (including discriminatory behavior), without assessing deeper (and possibly context specific) characteristics such as risk aversion, experience and implicit biases.

Overall, our results are mostly in line with models of implicit gender discrimination. We find that manipulating the information that loan officers have access to does not affect their lending behavior in a gender-specific way; that loan officer heterogeneity in implicit gender bias correlates with discriminatory behavior; and that discriminatory lending decisions do not boost loan quality.

This conclusion has direct repercussions for the policy implications we can draw from our experiment. For instance, interventions that increase information or the reliability of information about loan applicants may not be of first order importance for addressing gender bias in our context. Instead, policies to mitigate the real-world impact of loan officers' implicit biases may be called for. This could include simply making sure that loan officers have sufficient time to evaluate loan applications. Subtle interventions like this may be more effective than explicit diversity training programs, which make gender differences more salient and can even generate a backlash (Bohnet, 2016). Banks could also set bank-wide or branch-wide goals for lending to women without a guarantor. Management could then hold those that deviate from this norm accountable through disclosure and/or comply-or-explain procedures. Other easily implementable steps include making successful female entrepreneurs more visible to loan officers, for instance by integrating them in a natural way in banks' internal communication and training programs. Measuring the relative effectiveness of such interventions to contain the negative impact of implicit gender bias among loan officers provides a fruitful area for future experimental research.

References

- Abadie, A., S. Athey, G.W. Imbens and J. Wooldridge (2017), When Should You Adjust Standard Errors for Clustering?, NBER Working Paper No. 24003.
- Aigner, D.J. and G.G. Cain (1977), Statistical Theories of Discrimination in Labor Markets, *Industrial and Labor Relations Review*, 175-187.
- Agarwal, S. and I. Ben-David (2018), Loan Prospecting and the Loss of Soft Information, *Journal of Financial Economics* 129(3), 608-628.

- Agnew, J., P. Balduzzi and A. Sunden (2003), Portfolio Choice and Trading in a Large 401(k) Plan, *American Economic Review* 93, 193-215.
- Alesina, A.F., F. Lotti and P.E. Mistrulli (2013), Do Women Pay More for Credit? Evidence from Italy, *Journal of the European Economic Association* 11(s1), 45-66.
- Altonji, J.G. and R.M. Blank (1999), Race and Gender in the Labor Market, *Handbook of Labor Economics* 3, 3143-3259.
- Altonji, J.G. and C. Pierret (2001), Employer Learning and Statistical Discrimination, *Quarterly Journal of Economics* 116(1), 313-350.
- Arrow, K. (1973), The Theory of Discrimination, In: O. Aschenfelter and A. Rees (Eds.), *Discrimination in Labor Markets*, Princeton University Press, Princeton.
- Asiedu, E., J.A. Freeman and A. Nti-Addae (2012), Access to Credit by Small Businesses: How Relevant Are Race, Ethnicity, and Gender? *American Economic Review Papers & Proceedings* 102(3), 532-537.
- Andersson, P., (2004), Does Experience Matter in Lending? A Process-Tracing Study on Experienced Loan Officers' and Novices' Decision Behavior, *Journal of Economic Psychology* 25(4), 471-492.
- Banerjee, A.V., T. Besley and T.W. Guinnane (1994), Thy Neighbour's Keeper: The Design of a Credit Cooperative with Theory and a Test", *Quarterly Journal of Economics* 109(2), 491-515.
- Bayard, K., J. Hellerstein, D. Neumark and K. Troske (2003), New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data, *Journal of Labor Economics* 21(4), 887-922.
- Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande and P. Topalova (2009), Powerful Women: Does Exposure Reduce Bias?, *Quarterly Journal of Economics* 124(4), 1497-1540.
- Beck, T., P. Behr and A. Guettler (2013), Gender and Banking: Are Women Better Loan Officers?, *Review of Finance* 17(4), 1279-1321.
- Beck, T., P. Behr and A. Madestam (2018), Sex and Credit: Is There a Gender Bias in Lending?, *Journal of Banking and Finance* 87, 380-396.
- Becker, G.S. (1957), *The Economics of Discrimination*, University of Chicago Press (Chicago).
- Bellucci, A., A. Borisov and A. Zazzaro (2010), Does Gender Matter in Bank-Firm Relationships? Evidence from Small Business Lending, *Journal of Banking & Finance* 34(12), 2968-2984.

- Berg, T., (2015), Playing the Devil's Advocate: The Causal Effect of Risk management on Loan Quality, *Review of Financial Studies* 28(12), 3367-3406.
- Bertrand, M., D. Chugh and S. Mullainathan (2005), Implicit Discrimination, *American Economic Review Papers & Proceedings* 95(2), 94-98.
- Bertrand, M. and E. Duflo (2017), Field Experiments on Discrimination, In: *Handbook of Field Experiments (Vol. 1)*, E. Duflo and A. Banerjee (Eds.), Ch. 8, Elsevier, North Holland.
- Bertrand, M. and S. Mullainathan (2004), Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, *American Economic Review* 94(4), 991-1013.
- Blanchard, L., B. Zhao and J. Yinger (2008), Do Lenders Discriminate Against Minority and Women Entrepreneurs?, *Journal of Urban Economics* 63(2), 467-97.
- Blanchflower, D.G., P.B. Levine and D.J. Zimmerman (2003), Discrimination in the Small-Business Credit Market, *Review of Economics and Statistics* 85(4), 930-943.
- Bohnet, I. (2016), *What Works. Gender Equality by Design*, Harvard University Press, Cambridge, MA.
- Bond, P. and A.S. Rai (2008), Cosigned or Group Loans, *Journal of Development Economics*, 85, 58-80.
- Carlana, M. (2019), Implicit Stereotypes: Evidence from Teachers' Gender Bias, *Quarterly Journal of Economics*, 134(3), 1163-1224.
- Carney, D.R., B. Nosek, A.G. Greenwald and M.R. Banaji (2007), The Implicit Association Test (IAT), In: *Encyclopedia of Social Psychology*, R. Baumeister and K. Vohs (Eds.), Thousand Oaks, CA, Sage.
- Cole, S., M. Kanz and L. Klapper (2015), Incentivizing Calculated Risk-taking: Evidence from an Experiment with Commercial Bank Loan Officers, *Journal of Finance* 70(2), 537-575.
- Cornell, B. and I. Welch (1996), Culture, Information, and Screening Discrimination, *Journal of Political Economy* 104(3), 542-571.
- Demirgüç-Kunt, A., T. Beck and P. Honohan (2008), *Finance for All? Policies and Pitfalls in Expanding Access*, World Bank Policy Research Report, World Bank, Washington, D.C.
- Demirgüç-Kunt, A., L. Klapper, D. Singer, S. Ansar and J. Hess (2018), *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*, Washington, D.C. World Bank.

- Duarte, J., S. Siegel and L. Young (2012), Trust and Credit: The Role of Appearance in Peer-to-peer Lending, *Review of Financial Studies* 25(8), 2455-2483.
- EBRD (2019), *Business Environment and Enterprise Performance Survey (BEEPS)*, <https://ebrd-beeps.com/>, accessed on 27 May 2019.
- Eckel, C.C. and P.J. Grossman (2008), Differences in the Economic Decisions of Men and Women: Experimental Evidence, In: *Handbook of Experimental Economics Results*, Vol. 1, C. Plott and V. Smith (Eds.), 509-19, New York, Elsevier.
- Gneezy, U. and A. Imas (2017), Lab in the Field: Measuring Preferences in the Wild, In: A. Banerjee and E. Duflo (eds.), *Handbook of Field Experiments*, Vol. 1, Chapter 10, 439-464, Elsevier: North Holland.
- Greenwald, A.G., D.E. McGhee and J.L.K. Schwartz (1998), Measuring Individual Differences in Implicit Cognition: The Implicit Association Test, *Journal of Personality and Social Psychology* 74(6), 1464-80.
- Greenwald, A.G., T.A. Poehlman, E. Uhlmann and M.R. Banaji (2009), Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity, *Journal of Personality and Social Psychology* 97, 17-41.
- Heilman, M.E. (1984), Information as a Deterrent against Sex Discrimination: The Effects of Applicant Sex and Information Type on Preliminary Employment Decisions, *Organizational Behavior and Human Performance* 33(2), 174-86.
- Hertzberg, A., J.M. Liberti and D. Paravisini (2010), Information and Incentives Inside the Firm: Evidence from Loan Officer Rotation, *Journal of Finance* 65(3), 795-828.
- Kaas, L. and C. Manger (2012), Ethnic Discrimination in Germany's Labour Market: A Field Experiment, *German Economic Review*, 13(1), 1-20
- Klapper, L., A.M. Munoz and S. Singh (2014), The Gender Gap in the Use of Financial Services in Turkey, mimeo.
- Lane, K.A., M.R. Banaji, B.A. Nosek and A.G. Greenwald (2007), *Understanding and Using the Implicit Association Test*, New York, Guilford.
- Munnell, A.H., G.M.B. Tootell, L.E. Browne and J. McEneaney (1996), Mortgage Lending in Boston: Interpreting HMDA Data, *American Economic Review* 86(1), 25-53.
- Neal, D.A. and W.R. Johnson (1996), The Role of Pre-market factors in Black-White Wage Differences, *Journal of Political Economy* 104(5), 869-895.
- Neumark, D. (2018), Experimental Research on Labor Market Discrimination, *Journal of Economic Literature* 56(3), 799-866.

- Ongena, S. and A. Popov (2016), Gender Bias and Credit Access, *Journal of Money, Credit, and Banking* 48(8), 1691-1724.
- Phelps, E.S. (1972), The Statistical Theory of Racism and Sexism, *American Economic Review* 62(4), 659-661.
- Qian, J., P. Strahan and Z. Yang (2015), The Impact of Incentives and Communication Costs on Information Production and Use: Evidence from Bank Lending, *Journal of Finance*, LXX(4): 1457-1493.
- Reuben, E., P. Sapienza and L. Zingales (2014), How Stereotypes Impair Women's Careers in Science, *Proceedings of the National Academy of Sciences*, 111(12): 4403-08.
- Rooth, D.O. (2010), Automatic Associations and Discrimination in Hiring: Real World Evidence, *Labour Economics*, 17(3), 523-534.
- Sunden, A.E. and B.J. Surette (1998), Gender Differences in the Allocation of Assets in Retirement Savings Plans, *American Economic Review* 88, 207-211.
- The Economist (2013), [Discrimination Abounds. Women are Being Excluded from Finance Across the Developing World](#), November 19.
- World Economic Forum (2018), [The Global Gender Gap Report 2018](#), Geneva.

Table 1: Summary statistics

	N	Mean	Median	Sd.	Min	Max
<hr/> Panel A: Participant characteristics <hr/>						
Participant is female	332	0.47	0.00	0.50	0	1
Participant experience (years)	324	4.99	4.00	3.89	0	19
Participant age (years)	321	37.30	36.00	5.84	26	53
Participant is supervisor	334	0.43	0.00	0.50	0	1
Participant risk aversion	333	4.11	4.00	1.37	1	6
Participant gender bias (IAT)	325	0.33	0.34	0.32	-0.93	1.00
<hr/> Panel B: Loan-file characteristics <hr/>						
Real life performing	100	0.50	0.5	0.50	0	1
Real life non-performing (NPL)	100	0.25	0	0.44	0	1
Real life declined	100	0.25	0	0.44	0	1
<hr/> Panel C: Decision characteristics <hr/>						
<i>First round</i>						
Rejection dummy	1,336	0.39	0.00	0.49	0	1
Guarantor dummy	814	0.27	0.00	0.44	0	1
Subjective repayment probability	1,329	60.11	70.00	30.81	0	100
<i>Second round</i>						
Rejection dummy	1,334	0.36	0.00	0.48	0	1
Guarantor dummy	860	0.27	0.00	0.44	0	1
Subjective repayment probability	1,324	61.48	70.00	30.41	0	100
<hr/> Panel D: Treatment characteristics (second round) <hr/>						
No subj.	1,334	0.34	0	0.47	0	1
No obj.	1,334	0.33	0	0.47	0	1

Notes: This table displays summary statistics for the variables used in the empirical analysis. Panel A summarizes the main characteristics of all participants who took part in the experiment. Panel B displays summary statistics for the 100 loan application files used in the experiment. Panel C displays summary statistics at the decision level (participant-file combination). Panel D displays summary statistics for the information treatments in the second round of the experiment.

Table 2: Applicant gender and loan rejection

Dependent variable: Rejection dummy				
	[1]	[2]	[3]	[4]
Female applicant	-0.013 (0.023)	-0.013 (0.023)	-0.010 (0.024)	-0.010 (0.024)
Participant is female	0.023 (0.023)	0.029 (0.023)	0.021 (0.024)	0.026 (0.024)
Participant experience (years)	-0.002 (0.004)	-0.003 (0.004)	-0.003 (0.004)	-0.003 (0.005)
Participant age (years)	-0.005* (0.003)	-0.005* (0.003)	-0.005* (0.003)	-0.005* (0.003)
Participant is supervisor	0.100*** (0.032)	0.101*** (0.032)	0.099*** (0.032)	0.100*** (0.032)
Participant risk aversion		-0.012 (0.010)		-0.012 (0.010)
Participant IAT score			-0.000 (0.044)	-0.003 (0.044)
Constant	0.552*** (0.098)	0.604*** (0.103)	0.553*** (0.101)	0.607*** (0.107)
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.014	0.015	0.014	0.015
N	1,272	1,272	1,240	1,240

Notes: The dependent variable is a *Rejection dummy* that equals ‘1’ if the participant declines the credit application and ‘0’ if the participant approves it. The sample is restricted to the first round of the experiment. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 3: Applicant gender and loan rejection: Participant heterogeneity

Dependent variable: Rejection dummy						
	Participant gender		Participant experience		Participant age	
	Female	Male	Below median	Above median	Below median	Above median
	[1]	[2]	[3]	[4]	[5]	[6]
Female applicant	-0.001 (0.040)	-0.026 (0.035)	-0.018 (0.039)	0.004 (0.030)	0.021 (0.034)	-0.017 (0.033)
R-squared	0.018	0.010	0.011	0.041	0.024	0.039
N	576	664	560	680	520	720
Wald test p -value	0.299		0.320		0.207	
	Participant position		Participant risk aversion		Participant gender bias	
	Officer	Supervisor	Below median	Above median	Below median	Above median
	[7]	[8]	[9]	[10]	[11]	[12]
Female applicant	-0.046 (0.030)	0.014 (0.039)	-0.020 (0.061)	-0.007 (0.031)	0.002 (0.039)	-0.036 (0.032)
R-squared	0.010	0.029	0.045	0.010	0.025	0.026
N	740	500	360	880	624	616
Wald test p -value	0.105		0.404		0.212	
Participant covariates	Yes	Yes	Yes	Yes	Yes	Yes
File FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is a *Rejection dummy* that equals ‘1’ if the participant declines the credit application and ‘0’ if the participant approves it. The sample is restricted to the first round of the experiment. When partitioning continuous variables the “Below median” sample corresponds to strictly below the median while the “Above median” sample corresponds to values at the median and above. For the *Participant risk aversion* variable, higher values indicate greater risk aversion so that participants with above median risk aversion are the most risk averse. *Participant gender bias* measures implicit gender bias based on an implicit association test (IAT). Higher IAT values indicate that participants associate men more with careers and women more with household tasks. The Wald test p -value corresponds to one-sided tests. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 4: Applicant gender and guarantor requirements

Dependent variable: Guarantor dummy				
	[1]	[2]	[3]	[4]
Female applicant	0.068** (0.029)	0.068** (0.029)	0.069** (0.030)	0.070** (0.030)
Participant is female	-0.026 (0.032)	-0.033 (0.031)	-0.020 (0.033)	-0.027 (0.032)
Participant experience (years)	0.002 (0.005)	0.003 (0.005)	0.003 (0.006)	0.003 (0.006)
Participant age (years)	0.002 (0.004)	0.002 (0.004)	0.001 (0.004)	0.002 (0.004)
Participant is supervisor	0.036 (0.042)	0.035 (0.042)	0.044 (0.043)	0.042 (0.043)
Participant risk aversion		0.014 (0.012)		0.015 (0.013)
Participant gender bias (IAT)			-0.038 (0.063)	-0.038 (0.062)
Constant	0.036 (0.120)	-0.031 (0.137)	0.065 (0.121)	-0.007 (0.138)
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.064	0.063	0.062	0.061
N	772	772	752	752

Notes: The dependent variable is a *Guarantor dummy* that equals ‘1’ if the participant approves the credit application but requests a guarantor and ‘0’ if the participant approves it without requesting a guarantor. The sample is restricted to the first round of the experiment. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 5: Applicant gender and guarantor requirements: Participant heterogeneity

Dependent variable: Guarantor dummy						
	Participant gender		Participant experience		Participant age	
	Female	Male	Below median	Above median	Below median	Above median
	[1]	[2]	[3]	[4]	[5]	[6]
Female applicant	0.082 (0.052)	0.078 (0.049)	0.106** (0.052)	0.032 (0.046)	0.121** (0.050)	0.013 (0.040)
R-squared	0.107	0.080	0.097	0.077	0.136	0.037
N	338	414	341	411	325	427
Wald test p -value	0.473		0.108		0.035	
	Participant position		Participant risk aversion		Participant gender bias	
	Officer	Supervisor	Below median	Above median	Below median	Above median
	[7]	[8]	[9]	[10]	[11]	[12]
Female applicant	0.130*** (0.038)	-0.022 (0.061)	0.067 (0.065)	0.087* (0.044)	0.022 (0.051)	0.119** (0.046)
R-squared	0.117	0.034	0.161	0.041	0.063	0.090
N	471	281	214	538	381	371
Wald test p -value	0.008		0.389		0.055	
Participant covariates	Yes	Yes	Yes	Yes	Yes	Yes
File FE	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is a *Guarantor dummy* that equals ‘1’ if the participant approves the credit application but requests a guarantor and ‘0’ if the participant approves it without requesting a guarantor. The sample is restricted to the first round of the experiment. When partitioning continuous variables the “Below median” sample corresponds to strictly below the median while the “Above median” sample corresponds to values at the median and above. For the *Participant risk aversion* variable, higher values indicate greater risk aversion so that participants with above median risk aversion are the most risk averse. *Participant gender bias* measures implicit gender bias based on an implicit association test (IAT). Higher IAT values indicate that participants associate men more with careers and women more with household tasks. The Wald test p -value corresponds to one-sided tests. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 6: Applicant gender, guarantor requirements, and real-life loan performance

Dependent variable: Guarantor dummy								
	All		Performing loans					
	Loan in real life		Participant gender		Participant experience		Participant age	
	Performing	NPL & Declined	Female	Male	Below median	Above median	Below median	Above median
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Female applicant	0.124*** (0.040)	-0.022 (0.047)	0.119** (0.057)	0.113 (0.071)	0.145** (0.063)	0.076 (0.057)	0.157** (0.069)	0.092* (0.049)
R-squared	0.083	0.064	0.139	0.114	0.132	0.100	0.170	0.063
N	449	303	207	242	208	241	201	248
Wald test <i>p</i> -value	0.008		0.466		0.175		0.196	
			Participant position		Participant risk aversion		Participant gender bias	
			Officer	Supervisor	Below median	Above median	Below median	Above median
			[9]	[10]	[11]	[12]	[13]	[14]
Female applicant			0.161*** (0.051)	0.035 (0.077)	0.059 (0.078)	0.137** (0.057)	0.102 (0.063)	0.161*** (0.053)
R-squared			0.174	0.062	0.243	0.059	0.105	0.130
N			280	169	122	327	217	232
Wald test <i>p</i> -value			0.060		0.171		0.207	
Participant covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
File FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

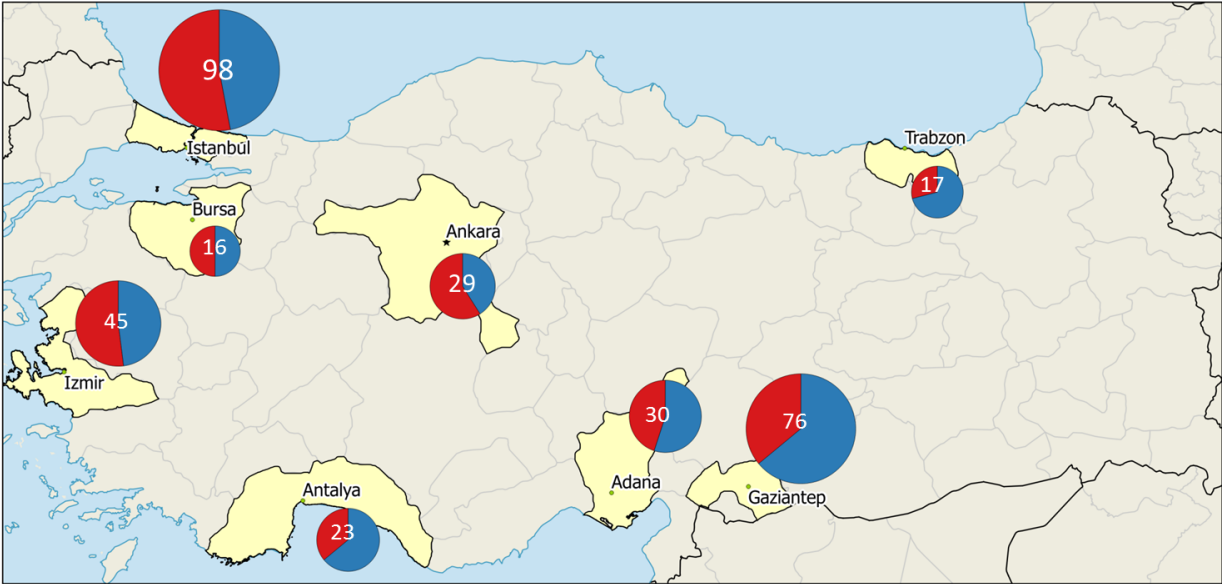
Notes: The dependent variable is a *Guarantor dummy* that equals ‘1’ if the participant approves the credit application but requests a guarantor and ‘0’ if the participant approves it without requesting a guarantor. The sample is restricted to the first round of the experiment. When partitioning continuous variables the “Below median” sample corresponds to strictly below the median while the “Above median” sample corresponds to values at the median and above. For the *Participant risk aversion* variable, higher values indicate greater risk aversion so that participants with above median risk aversion are the most risk averse. *Participant gender bias* measures implicit gender bias based on an implicit association test (IAT). Higher IAT values indicate that participants associate men more with careers and women more with household tasks. The Wald test *p*-value corresponds to one-sided tests. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table 7: Availability of borrower information and gender bias

Dependent variable:	Rejection dummy		Guarantor dummy	
	[1]	[2]	[3]	[4]
Female applicant	-0.017 (0.022)	0.014 (0.044)	0.035 (0.031)	0.023 (0.051)
No subj.	0.055* (0.029)	0.087** (0.041)	-0.031 (0.037)	-0.053 (0.050)
No obj.	-0.049 (0.036)	-0.035 (0.048)	-0.017 (0.033)	-0.014 (0.048)
No subj. × Female applicant		-0.064 (0.063)		0.043 (0.074)
No obj. × Female applicant		-0.028 (0.068)		-0.006 (0.074)
Participant covariates	Yes	Yes	Yes	Yes
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.036	0.037	0.066	0.066
N	1,238	1,238	802	802

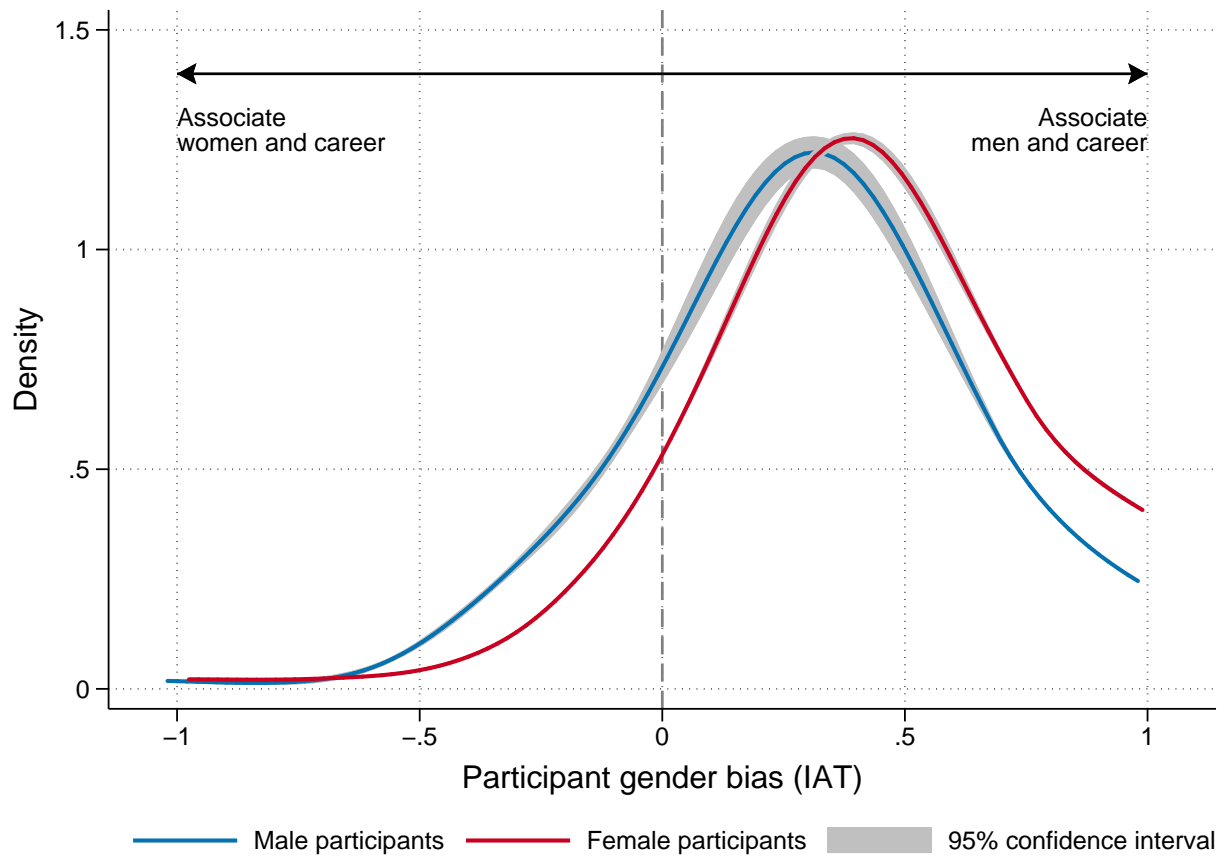
Notes: The dependent variable in columns [1] and [2] is a *Rejection dummy* that equals ‘1’ if the participant declines the credit application and ‘0’ if the participant approves it. The dependent variable in columns [3] and [4] is a *Guarantor dummy* that equals ‘1’ if the participant approves the credit application but requests a guarantor and ‘0’ if the participant approves it without requesting a guarantor. The sample is restricted to the second round of the experiment. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Figure 1: Geographical distribution of participants across the Turkish bank branches



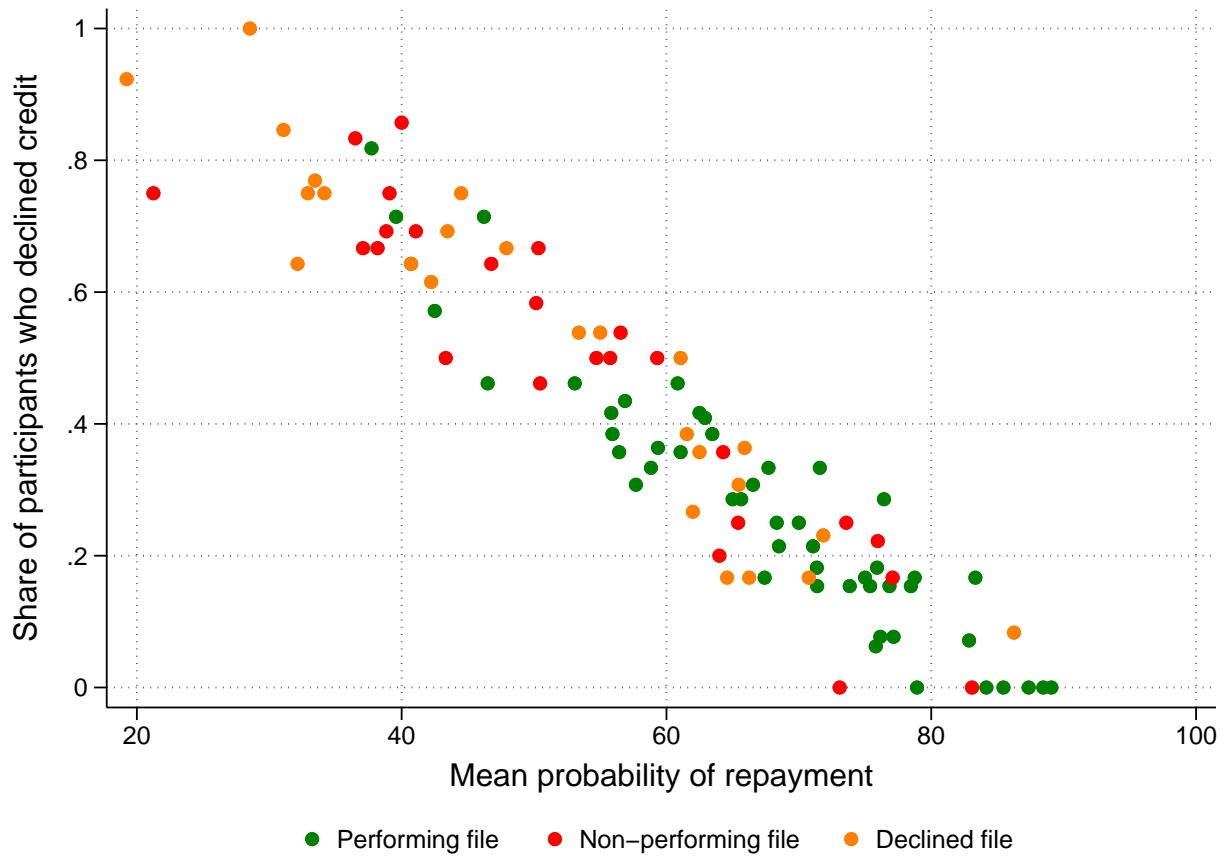
Notes: This map shows the number and gender composition of the participants in the eight Turkish regional bank branches that participated in the experiment. Circle size is proportional to the number of participants. The percentage of female (male) participants is shown in red (blue).

Figure 2: Participant gender bias (IAT), by participant sex



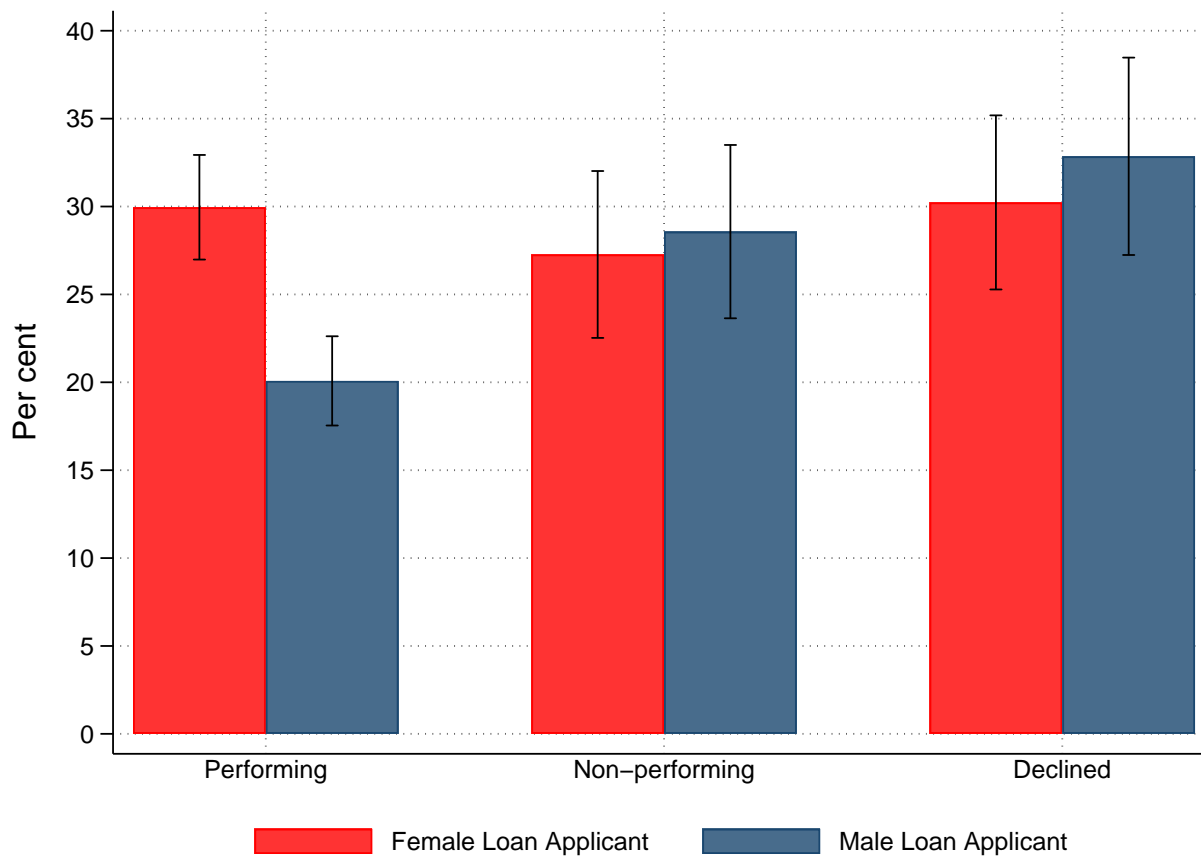
Notes: This figure shows a local polynomial smooth with 95 per cent confidence intervals of the variable *Participant gender bias (IAT)* for male (blue) and female (red) participants, respectively. The combined two-sample Kolmogorov-Smirnov test statistic is 0.181 and has a p-value of 0.01.

Figure 3: Expected repayment and loan rejection rates



Notes: The x-axis is the within-file mean, across participants, of the subjective repayment probability. The y-axis is the share of participants who declined the loan application. The figure is based on the first round of the experiment only.

Figure 4: Guarantor requirements, by loan quality and applicant sex



Notes: This figure shows the percentage of loan applications that were approved during the experiment and for which participants requested a guarantor. Separate bars are shown for approved loans that were repaid in real life (left), approved loans that were defaulted on in real life (middle), and loan applications that were rejected in real life (right). In each case, separate bars indicate applications that were shown to participants as coming from a female (red) or male (blue) entrepreneur. The whiskers indicate one binomial standard deviation. The sample is restricted to the first round of the experiment.

Appendices

Table A1: Variable definitions

Panel A: Participant characteristics	
Participant is female	Dummy variable equal to 1 for female and 0 for male participants.
Participant experience (years)	Number of years the participant has been an employee in the bank's credit division.
Participant age (years)	Age of the participant in years.
Participant is supervisor	Dummy variable equal to 1 for participants who are a supervisor/branch manager, 0 for those who are a loan officer.
Participant risk aversion	Integer variable ranging from 1 to 6, with 1 indicating risk loving and 6 indicating the highest level of risk aversion.
Participant gender bias (IAT)	Takes values from -1 to 1. Positive (negative) values indicate that the participant associates careers and entrepreneurship with being male (female). A score of zero indicates no implicit gender bias.
Panel B: File characteristics	
Female applicant	Dummy variable equal to 1 if the randomized gender of the loan application is female and 0 otherwise.
Female applicant (original)	Dummy variable equal to 1 if the gender of the real-life loan application was originally female and 0 otherwise.
Real life performing	Dummy variable equal to 1 if the loan was performing in real life, 0 otherwise.
Real life NPL	Dummy variable equal to 1 if the loan was non-performing in real life, 0 otherwise.
Real life declined	Dummy variable equal to 1 if the loan application was declined by the lending staff in real life, 0 otherwise.
Micro	Dummy variable equal to 1 if the credit file was from a micro firm and 0 if the credit file was from an SME firm.
Log of credit demanded	Logarithm of the amount of credit requested by the applicant.
Credit score	Credit score as taken from the KKB credit registry. Higher values indicate less ex ante credit risk.

Table A1 continued on next page

Table A1 continued

Panel C: Decision characteristics

Rejection dummy	Dummy variable equal to 1 if the participant rejects the loan application, 0 otherwise.
Guarantor dummy	Dummy variable equal to 1 if the participant offers credit conditional on the presence of a guarantor and 0 if the participant offers credit but does not request a guarantor.

Panel D: Treatment characteristics

No subj.	Dummy variable equal to 1 if information subjectively provided by lending staff is removed from the loan application file, 0 otherwise.
No obj.	Dummy variable equal to 1 if objective information (the credit score) from the credit bureau is removed from the loan application file, 0 otherwise.

Table A2: Applicant gender and credit score

Dependent variable: Credit score					
	[1]	[2]	[3]	[4]	[5]
Female applicant (original)	-12.845 (49.441)	51.042 (67.354)	59.297 (67.639)	66.736 (67.332)	79.874 (67.102)
Micro				-136.459* (70.387)	-39.468 (96.173)
Log of credit demand					68.671* (36.547)
Constant	1,035.730*** (29.942)	1,065.000*** (0.000)	964.336*** (138.865)	1,115.907*** (158.469)	299.568 (486.487)
Sector FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
R-squared	0.000	0.212	0.233	0.250	0.273
N	243	243	243	243	243

Notes: The dependent variable is *Credit score* (taken from the KKB credit registry) where higher values indicate less ex ante credit risk. The sample includes the 250 loan files from which the 100 loan files used in the experiment were drawn. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table A3: Gender, credit access, and firm growth

Dependent variable: Sales growth			
	[1]	[2]	[3]
Constrained female vs unconstrained male	-0.055 (0.197)	-0.103 (0.080)	-0.059 (0.062)
Industry FE	Yes	Yes	Yes
Year FE	No	Yes	Yes
Firm controls	Yes	Yes	Yes
R-squared	0.185	0.269	0.190
N	145	310	501

Notes: This table reports OLS regression results. The dependent variable is *Sales growth* which is the firm-level change in sales over the past three fiscal years. *Sales growth* is normalized between -1 and 1. *Constrained female vs unconstrained male* is a dummy variable equal to 1 if the principal owner (or one of the principal owners) of the firm is female and does not have a bank loan because she was rejected when applying for a loan at a bank or because she was discouraged from applying in the first place. This dummy is 0 if the principal owner (or one of the principal owners) is a male whose loan application was approved. In columns [1] and [2] ‘Firm controls’ include firm *Size*, *Originally private*, *Audited*, *Exporter* and *Competition*. In column [3] ‘Firm controls’ do not include *Competition* due to variable availability. Data are taken from the EBRD-World Bank Business Environment and Performance Survey (BEEPS), a face-to-face firm-level survey, for Turkey. The sample is restricted to BEEPS III in column [1]; BEEPS III and IV (pooled) in column [2]; and BEEPS III, IV and V (pooled) in column [3]. BEEPS III took place in 2005, BEEPS IV in 2009-09 and BEEPS V in 2012-14. Standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table A4: Correlation matrix

	Participant is supervisor	Participant is female	Participant age (years)	Participant risk aversion	Participant experience (years)	Participant gender bias (IAT)	Female applicant	Rejection dummy
Participant is supervisor	1.000							
Participant is female	0.091**	1.000						
Participant age (years)	0.570***	0.032	1.000					
Participant risk aversion	0.027	0.153***	-0.007	1.000				
Participant experience (years)	0.360***	0.031	0.503***	0.026	1.000			
Participant gender bias (IAT)	0.093**	0.196***	0.086**	-0.007	0.014	1.000		
Female applicant	0.000	0.000	-0.000	0.000	-0.000	0.000	1.000	
Rejection dummy	0.075**	0.037	0.014	-0.013	-0.015	0.008	-0.017	1.000

Notes: The sample is restricted to the first round. *, **, *** indicate significance at the 10, 5 and 1 per cent level, respectively.

Table A5: Predictors of participant gender bias

Dependent variable: Participant gender bias (IAT)	
	[1]
Participant is female	0.124*** (0.036)
Participant experience (years)	-0.004 (0.005)
Participant age (years)	0.004 (0.004)
Participant is supervisor	0.033 (0.044)
Participant risk aversion	-0.009 (0.014)
Constant	0.164 (0.146)
R-squared	0.049
N	310

Notes: The dependent variable is *Participant gender bias (IAT)* which takes values from -1 to 1. Positive (negative) values indicate that the participant associates careers and entrepreneurship with being male (female). A score of zero indicates no implicit gender bias. The sample is restricted to the first round round of the experiment. Standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.

Table A6: Gender and loan rejection: The role of supervisors

Dependent variable: Rejection dummy				
	[1]	[2]	[3]	[4]
Female applicant	-0.041 (0.029)	-0.041 (0.029)	-0.039 (0.029)	-0.039 (0.029)
Participant is female	0.024 (0.023)	0.029 (0.023)	0.021 (0.024)	0.027 (0.024)
Participant experience (years)	-0.003 (0.004)	-0.003 (0.004)	-0.003 (0.004)	-0.003 (0.005)
Participant age (years)	-0.005* (0.003)	-0.005* (0.003)	-0.005* (0.003)	-0.005* (0.003)
Participant is supervisor	0.067* (0.039)	0.067* (0.040)	0.064 (0.040)	0.065 (0.040)
Female applicant \times Participant is supervisor	0.067 (0.044)	0.068 (0.045)	0.070 (0.046)	0.071 (0.046)
Participant risk aversion		-0.012 (0.010)		-0.012 (0.010)
Participant IAT score			-0.000 (0.044)	-0.003 (0.044)
Constant	0.564*** (0.099)	0.617*** (0.105)	0.566*** (0.102)	0.620*** (0.108)
File FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
R-squared	0.016	0.017	0.016	0.017
N	1,272	1,272	1,240	1,240

Notes: The dependent variable is a *Rejection dummy* that equals '1' if the participant declines the loan application and '0' if the participant approves it. The sample is restricted to the first round of the experiment. Robust standard errors are in parentheses. *, **, *** indicate significance at the 10, 5, and 1 per cent level, respectively.