American Economic Journal: Applied Economics 2023, 15(2): 31–68
https://doi.org/10.1257/app.20210180

Discriminatory Lending: Evidence from Bankers in the Lab[†]

By J. Michelle Brock and Ralph De Haas*

We implement a lab-in-the-field experiment with 334 Turkish loan officers to document gender discrimination in small business lending and unpack mechanisms. Officers review multiple real-life loan applications in which we randomize applicant gender. While unconditional approval rates are the same, officers are 26 percent more likely to require a guarantor when we present the same application as coming from a female instead of a male entrepreneur. A causal forest algorithm to estimate heterogeneous treatment effects reveals that discrimination is concentrated among young, inexperienced, and gender-biased officers. Discrimination mainly affects female loan applicants in male-dominated industries, indicating how financial frictions can perpetuate entrepreneurial gender segregation across sectors. (*JEL* C93, G21, G32, J16, L25, L26, O16)

Across the world, female entrepreneurs borrow less from banks than male entrepreneurs do (Demirgüç-Kunt et al. 2018). Whether this gender gap is inefficient depends on whether it reflects differences in the demand for or the supply of loans. On the demand side, women may select into less capital-intensive firms that require little credit (Demirgüç-Kunt, Beck, and Honohan 2008). On the supply side, discrimination by lenders is often cited as contributing to women's financial exclusion (OECD 2016). In the latter case, female entrepreneurs face excessively

* Brock: European Bank for Reconstruction and Development, CEPR (email: brockm@ebrd.com); De Haas: European Bank for Reconstruction and Development, CEPR, and KU Leuven (email: dehaasr@ebrd.com). Seema Jayachandran was coeditor for this article. We thank Francesca Dalla Pozza, Gözde Esen, Lucas Kitzmüller, Melanie Koch, and Victoria Marino for excellent research assistance and Martin Acht, Sofia Amaral, Thorsten Beck, Tobias Berg, Miriam Bruhn, Zoe B. Cullen, François Derrien, Guido Friebel, Selim Gulesci, Sergei Guriev, Leora Klapper, Elif Kubilay, Sheisna Kulkarni, Lakshmi Naaraayanan, Sandra Orozco-Aleman, William Parienté, José-Luis Peydró, Alexander Popov, Margarita Tsoutsoura, Siddharth Vij, Luana Zaccaria, three anonymous referees, and seminar participants at the London School of Economics, King's College London, University of Oxford, Goethe University, Graduate Institute (Geneva), University of East Anglia, Ghent University, Copenhagen Business School, University of Southampton, University of Edinburgh, ZEW Mannheim, Higher School of Economics (Moscow), Durham University, KU Leuven, 2020 NBER Summer Institute, 2020 European Finance Association (EFA) Annual Meeting, Swiss Winter Conference on Financial Intermediation (Lenzerheide), Southern Economic Association Annual Meetings, eighth EFI Network Workshop, fourth EBC Network Workshop in Banking and Corporate Finance, MoFiR Virtual Seminar on Banking, second London School of Economics Workshop on the Political Economy of Turkey, sixth IWH-Fin-Fire Workshop, FDIC Center for Financial Research Seminar, the 2020 Fixed Income and Financial Institutions Conference (Darla Moore School of Business), and the third Endless Summer Conference on Financial Intermediation and Corporate Finance for useful comments. The views expressed are the authors' and not necessarily those of the institutions they are affiliated with. IRB approval was obtained from the Heartland Institutional Review Board (project 160920-23). The study is registered with the American Economic Association registry for Randomized Controlled Trials (RCT ID AEARCTR-0009025).

[†] Go to https://doi.org/10.1257/app.20210180 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

tight credit constraints that can leave productive capacity underutilized and that may ultimately hamper economic growth (Hsieh et al. 2019).

Discrimination in small business lending occurs when loan officers treat male and female applicants differently even if they are equal in all business-related aspects. Loan officers may hold female applicants to a higher standard either by directly rejecting women who do not meet this standard or by applying onerous conditions that make credit unattainable. Such indirect discrimination is particularly difficult to detect empirically. To test for the presence of both direct and indirect gender discrimination in small business lending, we implement a lab-in-the-field experiment in which loan officers evaluate multiple real-life loan applications where the gender of the applicant has been randomly manipulated by us. Bringing loan officers into a controlled environment allows us to carefully track their decisions and trace the mechanisms through which gender discrimination materializes.

We conduct our experiment with 334 loan officers of a large Turkish bank. Turkey provides a particularly suitable setting to study gender discrimination in lending. It is a large emerging market with a competitive banking system. The country scores well in terms of de jure gender equality: Few legal obstacles restrict women's ability to become an entrepreneur (Klapper and Singh 2014). At the same time, the country remains characterized by conservative gender norms. It only ranks at 130 out of 149 countries in terms of de facto gender equality (WEF 2018). This tension between gender-related laws on the book and actual attitudes within society characterizes many other emerging markets too.

We start by testing whether loan officers discriminate directly against female applicants. We find no evidence for such outright discrimination (although we may be underpowered to detect very small effects). Unconditional loan approval rates are similar when we present the same application as coming from a male or a female entrepreneur. We next investigate whether loan officers discriminate in a less direct way. We find strong evidence that they do. Loan officers are 26 percent more likely to make final loan approval conditional on the presence of a guarantor when we present the same application as coming from a female instead of a male entrepreneur. Because we use real-life loan applications that our partner bank received in the recent past, we can trace how loans performed in reality. We find that discrimination is concentrated among loans that were fully repaid in real life, making lending biases potentially costly to the bank.

We next investigate whether discrimination is widespread across the loan officer population or concentrated among certain types. We first estimate CATEs using sample-split and interacted regressions. We then apply machine learning—Wager and Athey's (2018) causal forest estimator—to more flexibly explore heterogeneous impacts. The algorithm identifies who discriminates most by predicting individual treatment effects based on loan officer traits. We find that younger and less experienced officers, especially those with stronger implicit biases against entrepreneurial women (measured via an implicit association test (IAT)), are more likely to impose discriminatory guarantor requirements.

We proceed by exploring two mechanisms that may underpin our results: gender differences in (actual or perceived) credit risk and loan officers acting on implicit biases informed by social norms. We find no evidence for the idea that loan officers are concerned about higher credit risk among female entrepreneurs. For example, the distribution of credit scores across male and female applicants is very similar, and loan officers themselves do not perceive female entrepreneurs to be riskier than equivalent male ones.

The second mechanism concerns implicit biases that may reflect social stereotypes (Bordalo et al. 2019). Our finding that loan officers with stronger implicit biases against entrepreneurial women are more likely to discriminate in terms of guarantor requirements already suggests that this mechanism plays an important role. To dig deeper, we divide our loan applications into those in relatively male-dominated versus female-dominated industries. We find that in stereotypically male industries, but not in female industries, loan approval is 10 percentage points more likely to be made conditional on a guarantor when we present the application as coming from a female instead of a male entrepreneur.

We again grow a causal forest to learn about treatment effect heterogeneity. The algorithm helps to disentangle the roles of loan officers' implicit gender bias, age, and work experience. We find that these moderators play distinct roles depending on whether women apply for a loan in a male or a female industry. In female-dominated industries, individual predicted treatment effects range between -2.7 and 10.6 percentage points. The algorithm reveals a tight negative relationship between loan officers' age and work experience and the predicted treatment effect on guarantor requirements. Once officers reach an age of about 45 (or, equivalently, just over 2 decades of work experience), they no longer discriminate against female applicants—that is, as long as women stick to traditionally female industries. In sharp contrast, when women apply for credit in gender-incongruent sectors, age and experience do not attenuate discrimination. Here, the predicted treatment effects are generally above 10 percentage points, and we find a tight positive link between the strength of officers' implicit biases and their predicted treatment effect. In sum, implicit biases underpin discriminatory guarantor requirements but do so in a context-specific way.

Our results advance the literature on several fronts. First, we address a gap in the literature on gender discrimination in entrepreneurial finance. Work using administrative data (Ewens and Townsend 2020) and experiments (Brooks et al. 2014) documents an investor bias against female entrepreneurs in need of venture capital. Hébert (2020), using French administrative data, shows that this equity funding gap reverses in female-dominated industries. There is less work on gender discrimination in entrepreneurial lending, and most of it relies on observational data.[1] Analyzing loans from an Italian bank, Bellucci, Borisov, and Zazzaro (2010) show that women 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. Women also need to post a guarantee more often. Similar studies from the United States find no gender discrimination.[2]

[1] Two recent papers focus on discrimination in consumer lending. Dobbie et al. (2021) use administrative data from a UK lender and find evidence for discrimination against immigrants and older applicants (but not women) due to an incentive scheme that biases loan officers against illiquid applicants. Montoya et al. (2020) randomly match stylized loan requests to male and female individuals who then apply by email for a small consumer loan. Requests submitted by women are less likely to be approved.

[2] See Blanchflower, Levine, and Zimmerman (2003) and Asiedu, Freeman, and Nti-Addae (2012).

We build on these papers by bringing loan officers to the lab and measuring traits that are typically unobservable—including implicit gender bias, risk preferences, and work experience. Employing recent advancements in causal machine learning, we show that some of these characteristics are first-order determinants of biased lending. This sheds new light on work by Beck, Behr, and Guettler (2013) and Beck, Behr, and Madestam (2018), who use data from an Albanian lender. The first paper shows that lending decisions by female loan officers result in fewer arrears, while the second finds that borrowers matched with opposite-sex loan officers pay higher interest rates. The first paper concludes 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" (Beck, Behr, and Guettler 2013, 5). Yet it acknowledges that performance differences between male and female loan officers may in fact reflect unobserved characteristics. We provide evidence to this effect by measuring such characteristics and quantifying their relative importance. Our causal forest shows that loan officers' implicit gender bias and work experience are seven and three times, respectively, more important than their own gender as drivers of discriminatory guarantor requirements.

Our experimental approach also reduces some identification concerns inherent to observational studies. In particular, we need not worry about omitted variables bias since we vary applicant gender while keeping all other characteristics of applications equal. We can also cleanly isolate the supply side of the credit market. This is important because a lower use of credit by female enterprises may simply reflect lower demand. Lastly, in administrative data, clients are typically not randomly matched to loan officers, which can bias estimates of discrimination. We instead randomly assign applications to loan officers so that there is no endogenous matching.

Second, we contribute to work investigating the drivers of discriminatory behavior. Ewens and Townsend (2020) propose a taxonomy of discrimination that distinguishes two broad categories. In the first one, discrimination is an efficient statistical process in which unbiased decision-makers use a group attribute as a signal of unobserved individual quality. For example, loan officers may know that the creditworthiness of men and women differs on average (Phelps 1972; Arrow 1973) or has a different variance (Aigner and Cain 1977). This can be referred to as accurate statistical discrimination. The second category comprises any kind of biased decision-making. Here, one can distinguish between taste-based discrimination (Becker 1957), inaccurate statistical discrimination (Bohren, Imas, and Rosenberg 2019; Bohren et al. forthcoming), and discrimination due to implicit biases (Neumark 2018). Taste-based discrimination occurs when decision-makers (say, loan officers) are prejudiced against a group (say, women) and avoid interacting with them or treat them unfavorably due to animus. Inaccurate statistical discrimination takes place when decision-makers hold miscalibrated beliefs about some outcome distributions (say, credit risk) across groups.[3] Discrimination due to implicit biases occurs when unconscious biases impact decision-making. Importantly, people are not always (fully) aware of their implicit biases (Bertrand, Chugh, and Mullainathan 2005).

[3] Miscalibrated beliefs can, for instance, take the form of gender stereotypes that contain a "kernel of truth" but exaggerate average differences (Bordalo et al. 2016, 2019).

Such biases may intensify taste-based discrimination, underpin inaccurate statistical discrimination, or directly influence decision-making.

To distinguish between different forms of discrimination, Bohren et al. (forthcoming) suggest to collect data on the subjective beliefs of evaluators. We do so by administering an IAT to measure loan officers' bias against entrepreneurial women.[4] Such bias may be most salient in male-centric domains (Reuben, Sapienza, and Zingales 2014). We indeed find that implicit biases have the strongest impact when women apply for a loan in a male-dominated sector. These results are at odds with models of accurate statistical discrimination and more in line with theories that highlight how implicit biases can affect decision-making.

Third, we contribute to research on the underrepresentation of women among entrepreneurs and on gender segregation across industries. For the United States, Gompers and Wang (2017) document that women constitute less than 10 percent of the entrepreneurial and venture capital labor pool. Women entrepreneurs also cluster in specific sectors, and this helps explain a large part of the gender wage gap (Blau and Kahn 2017). A separate strand of work explains the labor supply decisions of women and men as a function of deep-rooted social norms about the appropriate behavior of women (Alesina, Giuliano, and Nunn 2013; Grosjean and Khattar 2019) and men (Baranov, De Haas, and Grosjean forthcoming). These norms lead men and women to self-select into occupations that best match their self-perceived gender identity (Akerlof and Kranton 2010), forgo entrepreneurial opportunities at odds with prevailing norms (Field, Jayachandran, and Pande 2010), and be restricted in their choices because social norms have been codified into discriminatory laws (Naaraayanan 2020). Our contribution is to connect both lines of literature by showing how implicit biases about gender and entrepreneurship can generate financial frictions in the form of biased guarantor requirements, especially in traditionally male industries. Such frictions may then perpetuate an inefficient allocation of entrepreneurial talent across industries.

Lastly, our results add to a small literature on social collateral and third-party guarantees in lending. A guarantor takes legal responsibility for repayment in case the borrower fails to do so. Unlike passive collateral, guarantors actively monitor borrowers to ensure repayment, and monitoring is often leveraged by the threat of social sanctions (Bond and Rai 2008). This makes guarantees particularly effective in mitigating moral hazard (Pozzolo 2004).[5] The other side of the coin is that when a loan applicant is requested to ask a family member or friend to guarantee a loan, they put their social capital and reputation at risk. Guarantees thus tend to come at a social or psychological cost to the borrower. We show how implicit biases among loan officers expose female loan applicants considerably more to such guarantor requirements than otherwise identical male applicants. We also provide auxiliary, survey-based evidence indicating that many Turkish businesswomen perceive

[4] Attitude IATs measure implicit negative attitudes toward social groups. Stereotype IATs—like the one we use—measure implicit associations between social groups and specific traits (Bertrand and Duflo 2017).

[5] A related literature analyzes joint liability contracts in microfinance, where groups of (typically female) borrowers monitor each other, thus reducing moral hazard (Stiglitz 1990). Clients' dissatisfaction with the peer pressure in such group-based borrowing is a key reason behind the move toward individual liability microcredit over the past decades (Attanasio et al. 2015).

such biased guarantor requirements not only as unfair and costly, but also as a constraint on their ability to raise external finance.

The rest of the paper is structured as follows. Section I describes our setting and experimental design. Section II then summarizes the data generated by the experiment, outlines our estimation strategy, and introduces the causal forest algorithm. Section III presents the results, after which Section IV discusses mechanisms. Section V concludes.

I. Experimental Context and Design

A. *The Loan Approval Process*

We conducted our experiment in cooperation with a large commercial bank in Turkey. Over a 2-month period, 22 experimental sessions were held with a total of 334 bank employees across 8 cities.[6] The bank operates a regional office in each of these cities, and participants were randomly selected from all bank employees involved in small business lending (which makes up two-thirds of the bank's loan portfolio). Figure 1 shows the location of the regional offices and the number and gender of the participating bank employees.

Bank employees at two seniority levels participated in the experiment: loan officers (192) and supervisors (142). Both are located in branches and involved in the screening of borrowers. Loan officers establish contact with potential borrowers, conduct the initial screening, and collect documentation on business performance (income statements and balance sheets). They also check the availability of collateral and guarantors and request a credit score from the Turkish credit registry Kredi Kayıt Bürosu (KKB). Loan officers then enter this information into an electronic application form. They can also voluntarily add subjective notes to this form, such as notes about the client's perceived trustworthiness, 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 proposed maximum credit limit. Crucially, at this point, loan officers also recommend whether the loan application is approved unconditionally or made conditional on the presence of a guarantor. The supervisor then reviews the loan application and can reject or approve it. In the latter case, the application is sent to the bank's headquarters for formal sign-off.[7] Henceforth, we refer to the total experimental population as either "participants" or "loan officers."

[6] These were Adana, Ankara, Antalya, Bursa, Gaziantep, Istanbul, Izmir, and Trabzon. We also conducted a pilot session with 32 loan officers in Istanbul but do not use these pilot data.

[7] Branches can approve loans below a certain size threshold, but in practice only 10 percent of micro loans and 0.5 percent of loans to small and medium-sized enterprises (SMEs) are formally signed off in a branch. Micro clients are those with an annual turnover below TRY 2.5 million ($\approx$ US\$700,000) and a credit limit below TRY 750,000 ($\approx$ US\$210,000).

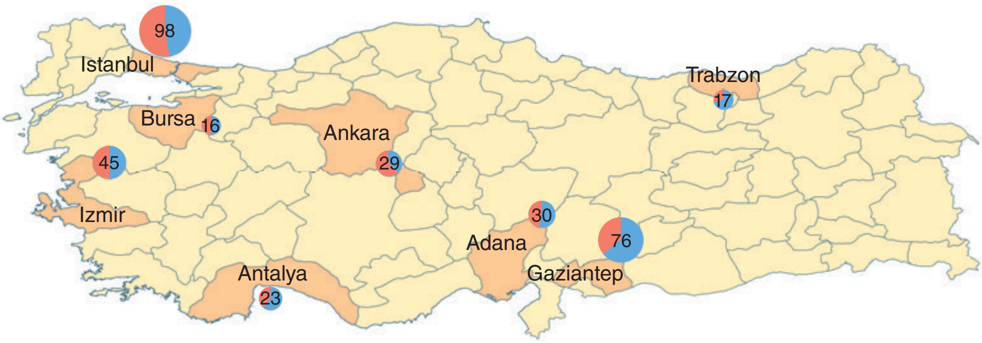

Figure 1. Geographical Distribution of Participants across the Bank's Regional Offices

Notes: This map shows the number and gender of the participants in the eight Turkish regional bank offices 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).

B. *Guarantor Requirements*

According to discussions with Turkish loan officers, the main function of requiring a guarantor is to leverage borrowers' social capital. Doing so attenuates ex ante moral hazard, thereby reducing the probability of default. If a borrower defaults nonetheless, banks can start legal proceedings against both the borrower and the guarantor simultaneously. In practice, however, loss given default of guaranteed versus nonguaranteed loans tends to be similar. One reason for this is that the legal process to recover a loan is lengthy.

It is important to understand whether guarantor requirements introduce an additional hurdle for loan applicants, especially female ones, and to what extent successful guarantor matches subsequently impose a social cost on borrowers. Guarantor requirements can impact applicants in two main ways. First, some entrepreneurs cannot find a guarantor, and their loan application may be denied as a result. Second, while other applicants may find a friend or family member willing to guarantee their loan, doing so puts their social capital on the line. Empirical evidence on these issues is scarce. For the case of Bangladesh, Singh, Asrani, and Ramaswamy (2016) and Jaim (2021) provide qualitative evidence that guarantor requirements are a significant barrier to female entrepreneurship. Similar evidence exists for Pakistan, where many female borrowers find it difficult to obtain a guarantor and, if successful, often have to pay guarantors a sizable amount as compensation (World Bank 2013). Consequently, many women remain cut off from bank loans. Recent experimental evidence from Vietnam (Diep-Nguyen and Dang 2020) shows that borrowers are willing to pay up to 9 percent of their monthly income to prevent repayment difficulties from being disclosed to their guarantor. This is consistent with the idea of high social costs associated with guarantors.

To the best of our knowledge, there exists no systematic evidence on how (much) guarantor requirements constrain female entrepreneurs in Turkey. To gain some insights into this, we conducted an online survey among a convenience sample of

Turkish businesswomen. The sample includes subscribers to *Business Lens*, a free online platform designed to provide women entrepreneurs with an assessment of their business' strengths and weaknesses. We fielded the survey in September 2021 using SurveyMonkey and received 208 fully or partially filled-out survey responses. Online Appendix C contains the survey instrument and summarizes all responses. Here, we only provide three key insights from the survey.

First, the survey indicates that Turkish businesswomen frequently encounter guarantor requirements. 61 percent of all respondents mention that they have ever been asked by a bank to provide a guarantor when they applied for any kind of loan or credit line. Among all those who applied for a *business* loan at any time in the past, 43 percent were asked to provide a guarantor as part of their most recent application. Moreover, 36 percent of all respondents have ever acted as a guarantor themselves and 54 percent of respondents think that banks are more likely to ask women for a guarantor than men.

Second, guarantor requirements impose financial constraints in practice. Forty-seven percent of respondents mention that a bank has at least once rejected their loan application because they could not provide a guarantor or did not want to provide one. When we ask respondents to rate, on a scale of 1 to 10, how difficult it is for an entrepreneur like them to find a guarantor, 39 percent of them pick 10. The average response is 7 out of 10.

Third, women entrepreneurs perceive guarantor requirements to be costly. In fact, 40 percent of all respondents are willing to pay a higher interest rate in order to get rid of the guarantor requirement.[8] One costly aspect of guarantor requirements is that they are perceived to be reciprocal. Almost half of all respondents (48 percent) believe that when someone agrees to act as their guarantor, there is "often" or "always" an expectation that they have to help them in some way in the future.

C. *Experimental Design*

Participants evaluated four applicant forms (henceforth, "loan applications") in the main part of the experiment.[9] We randomly presented these applications as coming from a woman or a man. 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 application, participants also had to provide a subjective repayment probability between 0 and 100. We did not constrain the time that participants had to evaluate the applications, and there was no feedback to participants about their decisions during the session. The sessions were framed as a general training exercise, and no gender-related issues were mentioned.

The task closely mimicked the daily choices that participants make in real life at work. Specifically, we presented all loan applications electronically and in the

[8] We provide the following scenario: "Suppose you want to take out a loan from a bank to finance an investment in your business that will cost 500,000 Turkish lira (for example, to pay for new machinery). The interest rate on this loan is 16 percent per year. The bank requires you to have a guarantor who co-signs the loan. Would you be willing to pay a higher annual interest rate in order not to have a guarantor?" On average, respondents are willing to pay a 5 percentage points higher interest rate.

[9] Participants made decisions on loan applications worth US$81.1 million in total.

standard application format that bank staff normally process on their computers. The loan applications contained all the information that was required for determining the creditworthiness of an applicant and that was available at the time the application was processed.[10] The loan applications did not include information about whether a guarantor was requested in real life.

We use 100 applications, selected from an initial sample of 250. These 250 applications were a stratified random sample of all first-time applications by existing SMEs (that is, no start-ups) that the bank received in the three to six years before the experiment.[11] Using this earlier period allows us to track what happened to each application in real life. The strata were region, gender, firm size, and whether the application was performing, nonperforming, or declined in real life. By using applications from applicants who had never before borrowed from our partner bank, we minimize the influence of soft information generated over time. All applications were gender neutral except for the randomly assigned name.

Each application was evaluated by 13.4 participants, half of the time as a female and half of the time as a male file. This allows us to obtain a within-application estimate of gender discrimination. Moreover, by asking participants to review both male and female applications, we preserve external validity, as no one at the bank sees only male or female clients. We indicate applicant gender by assigning new names, randomizing between male ones (Ahmet, Ali, Mehmet, Mustafa) and female ones (Ayse, Emine, Fatma, Zeynep). These names are common across Turkey and are well represented among working-age adults across regions.[12] No one saw the same file or same name more than once.

We held constant the ratio of performing, nonperforming, and rejected files that each participant saw, at 2–1–1. This ratio does not reflect the bank's actual application flow, but we used this ratio so that participants evaluated at least one file of each type. Names were randomized such that each participant saw one performing loan and one "bad" loan application (either a nonperforming loan or a declined application) from each gender.[13]

We incentivized decisions in line with common bank incentive schemes. Participants earned ten points (equivalent to ten Turkish lira) for each completed

[10]These forms are at the heart of the decision-making about whether the bank is willing to lend, what the maximum credit exposure will be, 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 means that during the experiment we could collect data on willingness to lend, maximum amount granted, and the need for a guarantor, but not on the interest rate or maturity of specific credit products. Online Appendix D contains a stylized loan application.

[11]When participants evaluated the files, they saw not the real application date but rather a date in the year of the experiment. We did so to avoid recall bias—loan officers did not have to think back about the economic situation in the past. This of course introduced a slight disconnect between loan performance in real life and the application evaluated during the experiment. To check whether this disconnect matters empirically, we regress our outcomes (loan rejection or guarantor requirement) on the difference between the loan application date and the time of the experiment, interacted with applicant gender. These interaction effects are never significant, indicating that the small timing difference does not have any gender-specific impact.

[12]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). When we include name fixed effects in our regressions, we fail to reject the null that these effects are jointly equal to zero.

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

review (quantity) and an additional five points when they correctly approved a loan that performed well in real life (quality).[14] 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 the incentive scheme at the bank, and the bank cannot realistically know when a rejection is incorrect.

Another way we replicate the real-world application process was to not incentivize guarantor decisions. The bank we worked with does not separately incentivize loan terms, including guarantor requirements. Officers hence request a guarantor when they expect it to increase the repayment probability without disproportionately increasing the risk that the applicant declines the offer. Incentivizing guarantor decisions would have entailed simulating the trade-off between repayment probability and the risk of a refusal. In principle, we could have done this by introducing a probability that the deal would fall through because of a guarantor request. This would have required both a more complicated setup and more file reviews per participant. The latter was particularly untenable. We therefore did not incentivize guarantor requirements, assuming that loan officers would rely on the heuristics they use in daily life to handle the abovementioned trade-off.

We aggregated all points per participant, and participants then exchanged points for prizes. Participants were ranked according to their score and split into four quartiles. In line with our instructions at the start of the session, those in the highest quartile could spend their points on 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 that they understood how the incentives worked and what the benefit would be of getting into the top quartiles. The incentive scheme was thus both material and competitive.

D. *Eliciting Personality Traits*

After the application decisions, we measured participants' risk preferences. We follow Eckel and Grossman (2008) and elicit risk preferences by presenting six risk scenarios, from which participants chose one. Each scenario was depicted as a circle split in half. Each half contained a possible outcome, in points, and the even split represented that the two outcomes were equally likely. The outcome pairs were 28–28, 20–44, 24–36, 16–52, 12–60, and 2–70. The task was incentivized: an on-site computer drew random draws to determine whether participants received the low or high number from the circle they selected.

Participants also took a stereotype IAT.[15] They had to sort, as quickly as possible, words that appeared sequentially on their tablet by clicking buttons at the right and

[14] This incentive scheme resembles the remuneration system that the bank uses in reality and is similar to the baseline scheme of Cole, Kanz, and Klapper (2015).

[15] IATs are by now common in psychology (Greenwald, McGhee, and Schwartz 1998) and economics (Bertrand, Chugh, and Mullainathan 2005; Glover, Pallais, and Parienté 2017; Carlana 2019). A meta-analysis 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).

left of the screen. The IAT started with two practice rounds in which participants sorted "career" words into a "career" bucket (left) and "family" words into a "family" bucket (right). This was repeated for male and female words.[16] After these practice rounds, the IAT mixed gender words and career/family words. Male and career words now shared a sorting button while female and family words shared the button on the other side of the screen (the stereotypical task). This was followed by another task where male and family words shared a sorting button while female and career words shared the other button (the nonstereotypical task). We recorded the time it took to sort each word in milliseconds. The assumption is that respondents with a stronger association between two concepts find sorting easier and complete it faster in one task compared to the other. We define a participant's implicit stereotype against entrepreneurial women as the normalized difference in mean response times between the nonstereotypical and the stereotypical task. Higher values indicate stronger stereotypes.[17]

An important design trade-off concerns the ordering of the application-review task and the IAT. Starting with the review task of male and female applications (as we did) might influence participants' subsequent IAT performance. Vice versa, starting with the IAT could prime participants about gender and hence affect subsequent lending decisions. We regard the former risk as much smaller for two reasons. First, the randomization of gender in the review task was subtle: participants had to work through four loan files with four different names (two male and two female). It is unlikely that this in itself would prime participants to think explicitly about gender. Second, in the review task, women were represented in equal proportion and with equal average quality. If anything, this could reduce stereotypical associations between men and career and between women and family. This might cause some downward pressure on the IAT score, but it is unlikely to impact the relative position of participants on the standardized scale. In contrast, the risk that participants would be primed to think about gender because of the IAT (which consists of male and female words appearing on their screen) would have been more acute.

II. Data and Estimation Strategy

A. *Data*

Table 2 summarizes our experimental data (Table 1 contains variable definitions). Panel A describes the characteristics of the 334 participants. Almost half of them are

[16] The IAT and all other documentation was provided in Turkish. The family-related words were translations for words such as "kitchen," "marriage," and "laundry." Career words included "office," "manager," and "job." To designate "male," we used words like "man," "boy," and "gentleman," and for "female" words, we used words such as "woman," "girl," and "lady."

[17] One may worry that IAT scores mainly proxy for cognitive ability. While raw IAT scores correlate with cognitive abilities (McFarland and Crouch 2002), this correlation is much weaker for standardized scores. Correlations with individual characteristics almost disappear when using a D-algorithm (Greenwald et al. 2003) instead of a raw score. We therefore use D-algorithm standardized IAT scores and correspondingly do not find much correlation between these scores and education level (a proxy for cognitive ability). The mean IAT score is 0.38 among those with secondary education or less, 0.33 among those with a Bachelor's or other postsecondary degree, and 0.32 among those with a Master's degree or PhD.

Table 1—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 of any 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	
Real life performing	Dummy variable equal to 1 if the loan was performing in real life, 0 otherwise.
Real life nonperforming (NPL)	Dummy variable equal to 1 if the loan was nonperforming 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.
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.
Credit score	Credit score as taken from the KKB credit registry. Higher values indicate less ex ante credit risk.
Credit limit requested (lira)	The total amount of credit requested by the applicant.
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.
Female-dominated sector	Dummy variable equal to 1 if the share of firms with majority female ownership, in a given industry, is greater than the median industry share; 0 otherwise. The share of female-owned firms is calculated at the two-digit ISIC level using pooled observations from the EBRD–World Bank BEEPS V and VI surveys.
Male-dominated sector	Dummy variable equal to 1 if the share of firms with majority female ownership in a given industry is less than or equal to the median industry share, 0 otherwise. The share of female-owned firms is calculated at the two-digit ISIC level using pooled observations from the EBRD–World Bank BEEPS V and VI surveys.
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.
Subjective repayment probability	Continuous variable that takes values from 0 to 100. For each decision, the participant estimates the likelihood that the loan would be repaid. Higher values indicate a greater chance of repayment.
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.

female, and their average age is 37 years. Forty-three percent of the participants are supervisors, the others are loan officers. There is substantial variation in the lending experience that loan officers have built up over the course of their career. While the

Table 2—Summary Statistics

	N	Mean	SD
Panel A. Participant characteristics			
Participant is female	332	0.47	0.50
Participant experience (years)	326	8.67	5.77
Participant age (years)	321	37.30	5.84
Participant is supervisor	334	0.43	0.50
Participant risk aversion	333	4.11	1.37
Participant gender bias (IAT)	325	0.33	0.32
Panel B. Loan-file characteristics			
Real life performing			
Female applicant (original)	50	0.66	0.48
Credit score	48	1,057	451
Credit limit requested (lira)	50	89,958	131,997
Female-dominated sector	49	0.73	0.45
Real life nonperforming (NPL)			
Female applicant (original)	25	0.32	0.48
Credit score	25	925	405
Credit limit requested (lira)	25	76,885	88,895
Female-dominated sector	24	0.71	0.46
Real life declined			
Female applicant (original)	25	0.40	0.50
Credit score	24	731	476
Credit limit requested (lira)	25	120,447	276,025
Female-dominated sector	23	0.74	0.45
Panel C. Decision characteristics			
First round			
Rejection dummy	1,336	0.39	0.49
Subjective repayment probability	1,329	60.11	30.81
Guarantor dummy	814	0.27	0.44

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). Table 1 contains all variable definitions.

average participant has worked as an officer for almost 9 years, this varies between less than 1 and 32 years.

Also summarized in Table 2 are results from the risk attitudes and IAT tasks, which leverage the lab-in-the-field setting and provide measures of participant characteristics that are otherwise difficult to observe. 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). Online Appendix Table A1 shows that this holds in our setting as well. The average risk aversion score is 4.32 (3.92) for women (men).

The IAT score is transformed so that it ranges between -1 and 1, with 0 indicating no implicit gender bias. While the scores vary widely, a large majority of lending staff (87 percent) have 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 (online Appendix Figure A1). The average IAT score is 0.39 for women and 0.28 for men, and this difference is statistically significant at the 5 percent level.[18]

Panel B of Table 2 summarizes the real-life characteristics of the 100 files. By design, half of these files refer to loans that in real life were paid back (performing), a quarter refer to loans that were defaulted upon (nonperforming), and another quarter are applications that were rejected in real life. As expected, credit scores were higher for loans that in real life performed well, as compared with either nonperforming loans or rejected applications. Just over 70 percent of the files are from sectors where female ownership is relatively common. As we discuss in the first subsection of Section IVB, we define female- and male-dominated sectors (at the two-digit International Standard Industrial Classification of All Economic Activities (ISIC) sector level) by the share of firms with majority female ownership. Female-dominated sectors are industries with an above-median share of female-owned firms.

Panel C summarizes the experimental outcomes at the participant-file decision level. Almost 40 percent of the loan applications are rejected outright, whereas, conditional on provisional acceptance, a guarantor is requested in 27 percent of the cases. For each application, we also asked the participant to estimate, on a 0–100 scale, the probability that the borrower would repay. The average estimated repayment probability is 60.1 percent.

These data also help 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 2, panel A provides a scatterplot of the 100 files. The horizontal axis indicates the average subjective repayment probability (each file was evaluated by 13.4 participants on average), while the vertical axis shows the share of participants that rejected the application in the lab. Figure 2 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 whether the decision-making in our lab in the field correlates with what happened to loan applications in real life. We find that this is the case. Overall, 72 percent of all applications that resulted in performing loans in real life were approved in the lab. This percentage is much lower for applications that resulted in nonperforming loans (53 percent) or were rejected in real life (47 percent).[19] As a result, files that in real life were nonperforming (gray dots) or declined (white) are concentrated in the upper-left corner of

[18] Online Appendix Table A2 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 helps explain 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 0.12 points (on the $[-1,1]$ scale) more biased against female entrepreneurs as compared with male bank employees.

[19] Online Appendix Figure A11 shows that in terms of initial approval decisions, there are no large differences between male and female applicants across all three types of applications. We do find, however, that for loan applications that in real life were declined, the probability of outright rejection in the experiment is 9 percentage points higher for women. This difference is borderline statistically significant (p-value $=$ 0.10).

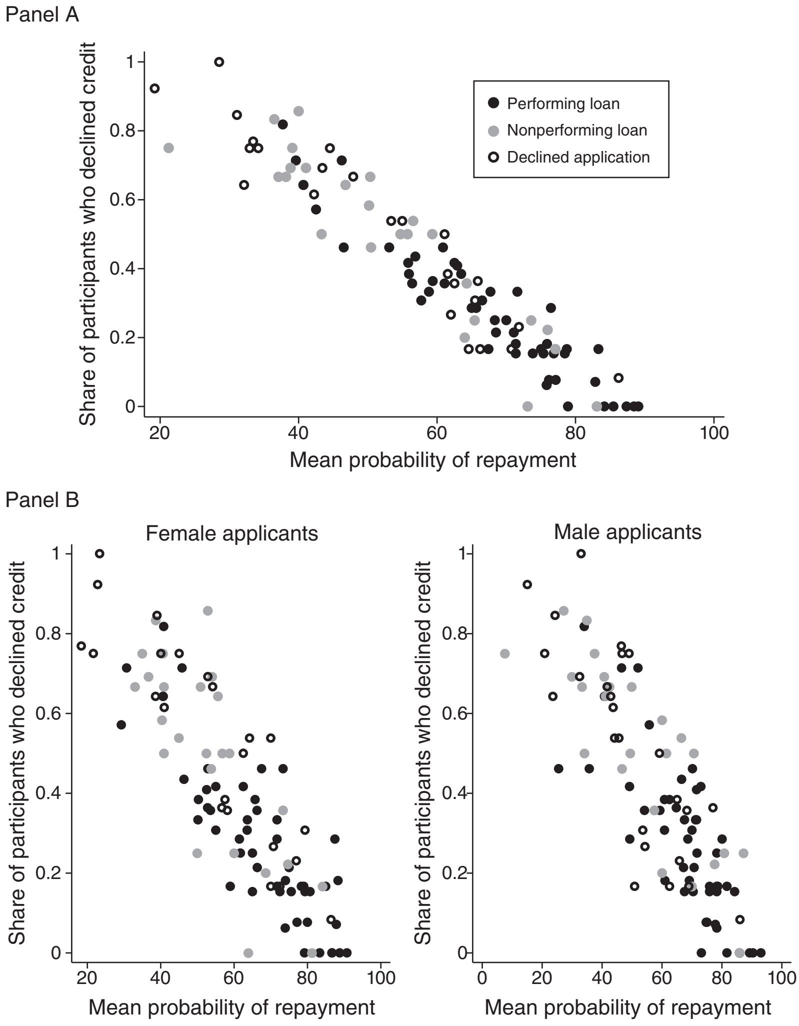

Figure 2. Expected Repayment and Loan Rejection Rates

Notes: The *x*-axis shows the within-file mean, across participants, of the subjective repayment probability. The *y*-axis shows the share of participants who declined the loan application. Panels A and B are based on the first round of the experiment; panel A corresponds to the full sample; and panel B splits the sample into two subsamples; applications from female (male) entrepreneurs are shown on the left-hand (right-hand) side of panel B. Table 1 contains all variable definitions.

Figure 2, panel A, while performing loans (black) are concentrated in the lower right-hand corner. Thus, across the board, participants correctly identified loans that performed well or badly in real life and made decisions in line with these subjective perceptions of loan quality. We obtain the same pattern independent of whether we present files as coming from a female (panel B, left) or a male applicant (panel B, right). This indicates that the loan officers were equally apt at identifying credit risk among male and female entrepreneurs.

B. *Estimation Strategy*

To test for biased lending behavior, we regress the application outcomes of interest, y_{il}, on G_{il}, the randomly assigned applicant gender of loan application l as seen by participant i. Our baseline specification is a parsimonious linear probability model with application (file) fixed effects, φ_l, which gives the within-file estimate of gender discrimination, β:

$$y_{il} = \alpha + \beta \cdot G_{il} + \varphi_l + \epsilon_{il}. \tag{1}$$

Standard errors, ϵ_{il}, are heteroscedasticity robust and clustered at the participant level. In all tables with subsample regression results, we also report Romano and Wolf (2005) step-down adjusted p-values, which control for the family-wise error rate and account for multiple hypothesis testing.[20]

Due to the experimental design, applicant gender is the only trait that varies across decisions about the same loan application. The application (file) fixed effects thus absorb all observed and unobserved file characteristics aside from applicant gender. Unobservables here include all (combinations of) features of the applications that the econometrician might ignore but that loan officers consciously or unconsciously care about. In this sense the experimental design and associated analytical specification provide stronger identification compared with observational studies where the data do not allow for within-file estimates.

An important question is whether we should saturate our baseline specification with additional covariates. If randomization was successful, our estimates of β will be unbiased. Online Appendix Tables A8–A11 provide balance tests that consistently show that participant traits are orthogonal not only to the treatment in the overall sample, but also in the various sample splits.[21] We therefore do not need covariates to arrive at unbiased estimates.

Even with successful randomization, covariates can improve precision and prevent tests on β from being underpowered. We therefore report two additional specifications. First, we add dummy variables for the city strata (where the experimental sessions took place). Second, we use double-LASSO regression, a disciplined way to let the data decide which participant covariates to include (if any) (Belloni et al. 2016, 2017).[22] In line with the successful randomization, LASSO in almost all cases tells us not to include covariates.

The successful randomization also obviates the need for participant fixed effects. This is important because we set up our experiment to arrive at a within-file (but between-participant) measure of possible discrimination by loan officers. Our

[20] We use Romano and Wolf's (2016) bootstrap resampling algorithm with 10,000 replications.

[21] In each table, the dependent variable is the *Female applicant* dummy (our treatment variable), which we regress on our six loan officer traits and file fixed effects. Across all regressions, most coefficients are close to zero and imprecisely estimated. As expected, some estimates are statistically significant, but there is no discernible pattern.

[22] We follow Belloni et al. (2016) to derive penalization parameters. Standard errors are cluster robust at the participant level (meaning that the penalty loadings account for heteroscedasticity and the clustered nature of the data). We estimate the double-LASSO within a fixed effect framework, which is equivalent to including unpenalized file dummies.

interest is in identifying how decision-makers judge the same loan file differently when we randomly present it as coming from a woman instead of a man. A limitation of this design—given the time constraints we had to work with—is that each officer could only review four loan files. In short, we did not set up the experiment in a way that would generate enough statistical power to include both file and participant fixed effects.

C. *Heterogeneous Treatment Effects*

Equation (1) provides estimates of the average treatment effect (ATE). We are also interested in conditional average treatment effects (CATEs) for subgroups of the loan officer population. In particular, we want to assess heterogeneity by loan officers' gender, work experience, age, position (junior loan officer versus supervisor), risk aversion, and implicit bias against entrepreneurial women. We follow two approaches. First, we present traditional sample-split regressions where we estimate equation (1) on subsamples (online Appendix Figure A3 also summarizes equivalent fully interacted regression models).

Second, we use supervised machine learning in the form of an honest causal forest algorithm to assess how impacts vary across loan officers (Athey and Imbens 2016; Wager and Athey 2018; Athey et al. 2019). Causal forests can combine multiple explanatory variables in a data-driven, nonlinear, and disciplined way. This gives us a more efficient, and hence statistically more powerful, tool to estimate heterogeneous treatment effects. Moreover, the algorithm tells us how useful each loan officer trait is in growing the forest. This allows us to gauge the relative importance of these traits as moderators of the causal effect between applicant gender and outcomes. We can also plot the value of these traits against the predicted treatment effect at the level of individual officers.

The algorithm grows a forest of causal trees. Each tree uses a random (bootstrapped) subsample of training data, the root node. The tree then recursively splits into increasingly smaller nodes that share similar covariates until it arrives at a set of terminal nodes (leaves). The algorithm makes splits that produce the biggest difference in treatment effects across leaves while still yielding an accurate estimate of the full treatment effect. If splitting a node does not result in an improved fit, that node is not split further and forms a leaf. This approach is honest in the sense that for each training subsample (that is, for each tree), observations are separated into a splitting sample (to determine where to place the splits) and an estimating sample (to estimate the within-leaf treatment effects).

We use the generalized random forest *grf* package for R by Tibshirani et al. (2020) to grow a forest of 20,000 trees based on a random training sample of 70 percent of the data. To grow each tree, we split the training sample into a splitting and estimating sample of equal size. This step is repeated 20,000 times. In a final step, the 30 percent of the dataset that was left aside is fed through all trees. For each one, we determine to which leaf each observation belongs based on the loan officer's traits. Each leaf indicates a specific predicted treatment effect—this is assigned to each observation associated with that leaf. The average prediction across all trees is then the predicted treatment effect at the officer level.

An alternative to honest causal forests for investigating heterogeneous treatment effects is generic machine learning inference, which is used to generate sorted group average treatment effects (GATES) (Chernozhukov et al. 2020). This approach is well suited for studies with rich baseline surveys and multiple ways of forming subgroups, increasing the risk of overfitting or selectively reporting results. In our case, however, the risk of overfitting is limited since we only have data on six baseline characteristics. We therefore systematically investigate potential heterogeneity along each of these six loan officer traits. A causal forest provides valid point-wise inference for CATEs when covariates are low dimensional, as in our case.[23] Using generic machine learning would entail a substantial efficiency loss because this approach accounts for not only sampling uncertainty due to estimation uncertainty regarding the parameter (conditional on the data split) but also uncertainty induced by the data splitting. This would be a high price to pay for addressing an issue (overfitting in case of numerous covariates) that is not particularly acute in our setting.

III. Results

A. *Applicant Gender and the Rejection of Loan Applications*

Table 3 presents linear probability regressions based on equation (1). The dependent variable is a *Rejection dummy*, which is "1" if an application was outright rejected by a participant and "0" if approved. The independent variable of interest, *Female applicant*, is a dummy for whether the application was presented as coming from a female ("1") or male ("0") entrepreneur. Column 1 shows a parsimonious specification with only file fixed effects, while column 2 adds city dummies as stratification controls. In column 3, we let double-LASSO pick from our six participant covariates as well as individual city dummies. As it turns out, columns 1 and 3 are identical because LASSO does not select any covariates.

Table 3 shows that we cannot reject the null hypothesis of no significant treatment effect of *Female applicant* on loan rejection. The coefficient for *Female applicant* is close to zero and, if anything, negative.[24] Since we include file fixed effects, our results show that the same application is not more likely to be outright rejected when we present it with a woman's name rather than a man's name. In short, we find no evidence of direct gender discrimination.

We also assess whether this null result applies to various subgroups. 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 standardized IAT score—and run sample-split regressions. We report these in online Appendix Table A3. There is no evidence of direct gender discrimination in any of these sample splits.

[23] According to Chernozhukov et al. (2020), a causal forest provides robust estimates if $\log(n) > d$, where n is the number of observations and d the number of dimensions of heterogeneity. In our case, $\log(814) = 6.7 > 6$.

[24] Our experiment was not powered to detect such a small effect, and the 95 percent confidence interval is therefore quite wide at $[-0.055, 0.040]$. To achieve 80 percent power to detect whether $\beta = -0.008$ is statistically nonzero would have required over 10,000 decisions—ten times our current sample.

TABLE 3—APPLICANT GENDER AND LOAN REJECTION

Dependent variable: Rejection dummy	(1)	(2)	(3)
Female applicant	−0.008	−0.008	−0.008
	(0.024)	(0.024)	(0.024)
R^2	0.259	0.264	0.259
Observations	1,336	1,336	1,336
File FE	✓	✓	✓
City FE		✓	
Double LASSO			✓

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. In column 3, a double-LASSO procedure is used to select controls from participant covariates and city FE (set of potential controls). The sample is restricted to the first round of the experiment. Cluster robust standard errors are shown in parentheses and clustered at the participant level. Table 1 contains all variable definitions.

While officers do not discriminate at the extensive margin (provisional approvals), they may do so at the intensive margin by providing women with smaller loans. As part of the (real-world) applications that officers reviewed in the lab, they saw the credit limit requested by the applicant. Conditional on initial approval, participants had to indicate whether they were willing to provide the full amount requested or less. In 60 percent of the cases, participants approved the full amount. When we regress the difference between the asked and the offered amount on *Female applicant*, the estimate is not statistically significant. The same holds when we simply regress the amount offered on this dummy while including file fixed effects. These results can be found in online Appendix Table A12.

B. *Applicant Gender and Guarantor Requirements*

We next test for a more indirect form of gender discrimination. In Table 4, we assess whether loan approval is more likely to be conditional on the presence of a guarantor when the application comes from a woman instead of a man, all else equal. We find strong evidence of such indirect discrimination: officers are 6 percentage points more likely to make final approval conditional on a guarantor when the application is shown as coming from a female instead of a male entrepreneur. The statistical and economic significance of this effect is stable across specifications.[25] The effect is large as only 27 percent of all preapproved applications are required to have a guarantor. This indirect discrimination implies that female entrepreneurs without a guarantor remain deprived of credit even if the officer in principle views the application favorably. To the extent that these entrepreneurs are in fact good credit risks, such a bias will be disadvantageous to the bank. Moreover, even for female borrowers who can provide a guarantor, putting their social capital on the line may be costly.

[25] In column 3, LASSO only picks one city dummy as a control. The results are robust to designating both the city fixed effects and *Participant is supervisor* (the other stratification variable) as unpenalized LASSO controls. We also obtain very similar results when including all six participant covariates or subsets of these.

Table 4—Applicant Gender and Guarantor Requirements

Dependent variable: Guarantor dummy	(1)	(2)	(3)
Female applicant	0.063	0.058	0.060
	(0.030)	(0.030)	(0.030)
R^2	0.152	0.188	0.173
Observations	814	814	814
File FE	✓	✓	✓
City FE		✓	
Double LASSO			✓
Better Lee Bounds		0.057, 0.061	
		[0.000, 0.118]	

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. In column 3, a double-LASSO procedure is used to select controls from participant covariates and city FE (set of potential controls). *Better Lee Bounds* refer to Lee (2009) bounds that are tightened through a LASSO selection procedure that considers all participant covariates (Semenova 2021). Stoye (2009)–adjusted Imbens and Manski (2004) 95 percent confidence intervals are reported in brackets below these bounds. The sample is restricted to the first round of the experiment. Cluster robust standard errors are shown in parentheses and clustered at the participant level. Table 1 contains all variable definitions.

Section IIIA already showed that, conditional on initial loan approval, female applicants do not benefit from larger loans. The stricter guarantor requirements imposed on them are hence not simply a quid pro quo for receiving more credit. We now assess more directly the link between loan amount granted and guarantor requirements. For each application that was provisionally approved, we take the difference between the amount demanded and approved by the loan officer. We standardize this difference as a z-score. When we correlate this z-score with a dummy for whether a guarantor was requested, the correlation coefficient is -0.01 overall as well as for male- and female-presented files separately. Differences between the amount asked and supplied are thus uncorrelated with the presence of a guarantor requirement. This holds equally for male and female applicants.

The regressions in Table 4 are based on fewer observations than those in Table 3 because the guarantor decision is conditional on initial loan approval.[26] To account for this selection, we provide Better Lee Bounds (Semenova 2020) below all guarantor regressions.[27] We report the lower and upper bounds as well as Stoye's (2009) version of the Imbens and Manski (2004) 95 percent confidence intervals.[28] Table 4

[26] Ninety-eight percent of the participants occur in both rejection and guarantor estimations, so there is no notable self-selection at the participant level.

[27] To construct bounds, we discretize age, experience, and IAT into quantiles and use the formula in Belloni et al. (2017) to set the LASSO penalty parameter. In the first stage, we estimate conditional selection by logistic LASSO equation and the conditional outcome equation by quantile LASSO. In the second stage, we plug estimates from the first stage into an orthogonalized moment equation (corrected for bias) for the bounds and report the sample average. We thank Vira Semenova for kindly sharing an updated version of her *leebounds* R package with us.

[28] We test the monotonicity assumption by comparing the difference in means between the treatment and control groups for each of the participant traits, using all observations that make it into the guarantor decision phase (cf. Lee 2009). These means are never statistically significantly different, with p-values ranging between 0.19 and 0.97. We carry out a joint significant test by regressing the treatment indicator (*Female applicant*) on the participant traits (using the same sample). The p-value of this F-statistic is 0.68 in Table 4. Online Appendix Tables A8–A11 present F-test p-values for all subsample guarantor regressions.

shows tight bounds for the treatment effect. To three decimal places, zero is just included in the 95 percent confidence interval of these bounds. Overall, it does not appear to be the case that the guarantor effect that we document is simply an artifact of selection into the guarantor stage based on the initial approval decisions.

We next assess the stability of our estimates across geographies and sectors. Panel A of online Appendix Figure A2 depicts coefficient estimates similar to those in column 1 of Table 4. Each estimate reflects a sample in which we drop observations from one city where a lab session took place (and where the participating loan officers are based). This visualizes how stable the results are across the experimental locations. We find that in all cases the coefficient indicates a 5 to 10 percentage point higher likelihood that a guarantor is requested from female applicants. The coefficients are ordered, from top to bottom, by decreasing average disposable household income in the excluded city. There is no apparent relationship between indirect gender discrimination and local economic development.

Panel B of online Appendix Figure A2 repeats this exercise, but now considering the region where each real-life application originated.[29] We drop one region at a time and plot the estimated coefficients, ordering them from the highest (top) to the lowest (bottom) regional income level per capita in 2016. We again find little geographic heterogeneity: in each case, the probability that a guarantor is required is between 5 and 7 percentage points higher when we present the same application as coming from a female rather than a male entrepreneur. Lastly, in panel C of Figure A2, we exclude one of the following macro sectors at a time: retail, services, manufacturing, wholesale, and other industries. The results again show a coefficient that consistently lies around 6 percentage points. We now assess whether biased guarantor requirements occur in the loan officer population as a whole or are instead concentrated among particular types of loan officers.

C. *Indirect Gender Discrimination: Participant Heterogeneity*

Heterogeneous Treatment Effects: Sample Splits.—Table 5 investigates heterogeneity in biased guarantor requirements through the lens of sample-split regressions. To follow a consistent approach as to which covariates (participant traits and/or city dummies) to include, we again use double-LASSO. As in Tables 3 and 4, in almost all cases, LASSO does not pick any covariates, with the exception of a city dummy in a few specifications and *Participant experience* in one specification. This signifies that there is balance not only of gender, but also in terms of the files used across cities and that participants were largely interchangeable between cities. We therefore present parsimonious specifications that only contain the file fixed effects—as in column 1 in Tables 3 and 4.[30]

We find a consistent pattern of CATEs. When we present the application as coming from a woman instead of a man, officers are more likely to ask for a guarantor

[29] The regions are Marmara, Aegean, Central Anatolia, Mediterranean, Black Sea, Eastern Anatolia, and Southeastern Anatolia.

[30] When we partition nonbinary 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.

Table 5—Applicant Gender and Guarantor Requirements: Participant Heterogeneity

Dependent variable: Guarantor dummy	Participant gender		Participant experience		Participant age	
	Female (1)	Male (2)	Below median (3)	Above median (4)	Below median (5)	Above median (6)
Female applicant	0.065	0.084	0.123	0.047	0.126	0.020
	(0.048)	(0.043)	(0.055)	(0.044)	(0.053)	(0.042)
	[0.421]	[0.107]	[0.051]	[0.508]	[0.042]	[0.843]
t-test p-values	0.387		0.139		0.060	
R^2	0.307	0.223	0.293	0.268	0.278	0.244
Observations	365	443	367	427	333	447
File FE	✓	✓	✓	✓	✓	✓
Better Lee Bounds	0.021, 0.099	0.028, 0.089	0.085, 0.090	0.018, 0.028	0.115, 0.143	0.003, 0.005
	[−0.080, 0.197]	[0.006, 0.102]	[0.004, 0.177]	[−0.063, 0.107]	[0.020, 0.224]	[−0.083, 0.096]
	Participant position		Participant risk aversion		Participant gender bias	
	Officer (7)	Supervisor (8)	Below median (9)	Above median (10)	Below median (11)	Above median (12)
Female applicant	0.120	−0.030	0.081	0.074	0.015	0.111
	(0.038)	(0.054)	(0.069)	(0.035)	(0.048)	(0.043)
	[0.003]	[0.843]	[0.507]	[0.075]	[0.843]	[0.018]
t-test p-values	0.012		0.466		0.067	
R^2	0.230	0.322	0.312	0.176	0.296	0.284
Observations	491	323	230	582	399	393
File FE	✓	✓	✓	✓	✓	✓
Better Lee Bounds	0.137, 0.145	−0.052, −0.052	—	0.053, 0.070	0.037, 0.042	0.074, 0.077
	[0.054, 0.205]	[−0.104, 0.080]	—	[0.000, 0.124]	[−0.042, 0.119]	[−0.009, 0.153]

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 nonbinary 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 IAT. Higher IAT values indicate that participants associate men more with careers and women more with household tasks. The t-test p-value corresponds to one-sided tests. Romano-Wolf p-values are shown in square brackets and refer to Romano-Wolf stepdown p-values that control for the family-wise error rate and account for multiple hypothesis testing; we adjust for 12 hypothesis and follow Romano and Wolf's (2016) bootstrap resampling algorithm with 10,000 replications. *Better Lee Bounds* refer to Lee (2009) bounds that are tightened through a LASSO selection procedure that considers all participant covariates (Semenova 2021). Stoye (2009)–adjusted Imbens and Manski (2004) 95 percent confidence intervals are reported in brackets below these bounds. Cluster robust standard errors are shown in parentheses and clustered at the participant level. Table 1 contains all variable definitions.

when they are younger (columns 5–6), in a more junior position (columns 7–8), and/or display more implicit gender bias in our IAT (columns 11–12). For example, officers with above-median levels of implicit gender bias are 11 percentage points more likely to request a guarantor when we present a file as coming from a female entrepreneur.[31] t-tests confirm that we can reject equality of coefficients in these pairs of βs at at least the 10 percent level.[32] There is also some evidence that participants with a below-median level of lending experience are more likely to ask women

[31] A few (42) loan officers display a negative gender bias, meaning that they associate women—rather than men—with a career. In line with symmetric interaction effects, we find that these officers are less likely to request a guarantor when we present an application as coming from a woman.

[32] We summarize results from equivalent fully interacted regression models in online Appendix Figure A3. The independent variables include *Female applicant*, an interaction of this dummy and a participant trait (such as *Participant experience*), and additional interactions between this trait and the file fixed effects. The bars show the coefficients for *Female applicant* and its interaction with the respective trait. The black dots indicate the sum of both coefficients.

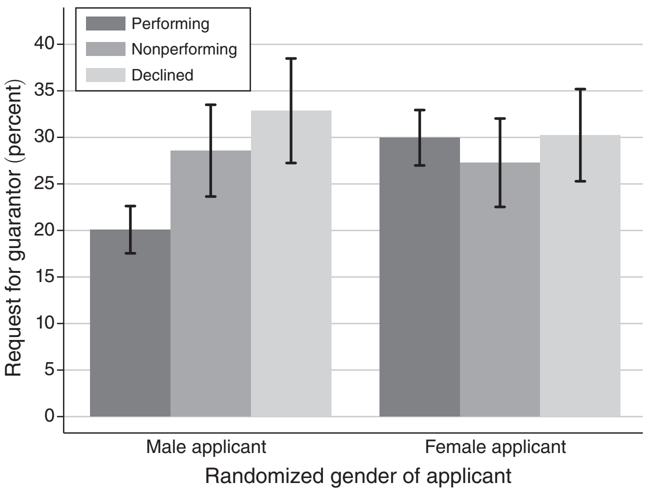

Figure 3. Guarantor Requirements, by Loan Quality and Applicant Gender

Notes: This figure shows the percentage of loan applications approved during the experiment and for which participants requested a guarantor. Bars are shown for approved loans repaid in real life (dark gray), approved loans that were defaulted on in real life (medium gray), and loan applications rejected in real life (light gray). Bars indicate applications that were shown to participants as coming from a female (right) or male (left) entrepreneur. Whiskers indicate one binomial standard error. The sample is restricted to the first round of the experiment. Table 1 contains all variable definitions.

for a guarantor (columns 3–4). These results suggest that age and seniority, possibly summarized by experience, reduce the extent to which officers use gender as a mental shortcut to determine whether a guarantor is required. Meanwhile, columns 1 and 2 of Table 5 show no significant difference between male and female participants in how they treat female applicants. There is also no significant difference between participants that are more or less risk averse (columns 9 and 10).[33]

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 now ask whether the higher probability that female loan applicants are required to have a guarantor is driven by loans that performed well in real life or by those that did less well. Figure 3 gives a nonparametric answer to this question. We divide all loan applications into those that were accepted in real life and performed well (dark gray bars), those that were accepted and became nonperforming (medium gray), and those that were declined in real life (light gray). The data pattern is striking. When we present

[33] There are no Better Lee Bounds in columns 9 and 11 of Tables 5 and 6, respectively. These subsamples lack sufficient variation in some variables to estimate the outcome equation by quantile LASSO and to move to the second stage of the procedure. Moreover, in a few instances (such as column 10), the coefficient estimate is just above the upper bound (though within the confidence interval). This can occur because for the bounds, we use all six participant covariates and let LASSO decide which ones matter for the selection and outcome equations. For the main regressions in Tables 5 and 6, we instead use a harmonized specification that only includes the treatment dummy and file fixed effects.

files as coming from male loan applicants (left-hand side), loan officers clearly and strongly differentiate between high-quality and lower-quality loans. For loans that were repaid in real life, men are asked for a guarantor in only 20.1 percent of the cases. This number is substantially higher for nonperforming loans and applications that were declined in real life, at 28.6 and 32.9 percent, respectively (these percentages are statistically different from that for performing loans with $p = 0.10$ and $p = 0.02$, respectively).

When we instead present the same files as coming from female loan applicants (right-hand side), the higher-quality loan applications do not benefit from lower guarantor requirements at all. It appears that women are held to a higher standard: even in the case of high-quality loan applications, there is still a 30 percent likelihood that a guarantor is requested. This is about the same percentage as for *low-quality* applications from male applicants. The data therefore show that it is among the better-quality loans that officers discriminate against female applicants. A similar picture emerges when we split the sample into applicants with an above- or below-median subjective repayment probability (online Appendix Figure A4, panel A) or into applicants with low, median, or high ex ante credit risk as measured by their credit score (online Appendix Figure A4, panel B). In both cases, gender discrimination in terms of requested guarantors is concentrated among applications with less ex ante credit risk.

In Table 6, we perform this analysis parametrically. Column 1 confirms that also when controlling for file fixed effects, women are 11.1 percentage points more likely to be asked for a guarantor in case of high-quality loans. This gender effect is absent for loans that were either rejected or nonperforming in real life (column 2), and this difference between high- and low-quality loan applications is statistically significant at the 5 percent level. This confirms that double standards are applied in the case of relatively good loans that were paid back in real life. Columns 3 to 14 reveal similar heterogeneity as before. High-quality female applications are 10 to 16 percentage points more likely to be asked for a guarantor compared to identical male applications if the participant is relatively inexperienced (columns 5–6), relatively young (columns 7–8), a loan officer rather than a supervisor (columns 9–10), and revealed a strong gender bias in our IAT (columns 13–14).[34] In summary, especially more junior and more gender-biased officers resort to the applicant's gender as a heuristic when there is no clear indication that a loan is risky.

Heterogeneous Treatment Effects: Honest Causal Forests.—The first subsection of Section IIIC provided a first analysis of CATEs. We now introduce a causal forest algorithm to more flexibly and efficiently disentangle how officer traits play distinct moderating roles in the causal relationship between applicant gender and guarantor requirements. Online Appendix Figure A5 (panel A) depicts the distribution of the predicted treatment effects. In the absence of treatment heterogeneity, the distribution would cluster tightly around the ATE of 6 percentage points. Instead, the causal

[34] Differences by participant gender (columns 3–4) and risk aversion (columns 11–2) are again smaller. Even where the subsample coefficients differ substantially, this difference is less precisely estimated due to the smaller sample (performing loans only). This is reflected in the *t*-test *p*-values at the bottom of Table 6.

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	0.111	−0.014	0.080	0.128	0.166	0.102	0.149	0.095
	(0.040)	(0.054)	(0.063)	(0.061)	(0.070)	(0.059)	(0.071)	(0.056)
	[0.001]	[0.736]	[0.302]	[0.079]	[0.036]	[0.171]	[0.079]	[0.171]
t-test p-values	0.031		0.292		0.245		0.274	
R^2	0.128	0.190	0.231	0.205	0.237	0.271	0.214	0.225
Observations	486	328	225	257	221	256	205	262
File FE	✓	✓	✓	✓	✓	✓	✓	✓
Better Lee Bounds	0.081, 0.143	−0.055, −0.007	0.038, 0.078	0.079, 0.218	0.079, 0.179	0.079, 0.083	0.150, 0.196	0.039, 0.076
	[0.082, 0.149]	[−0.176, 0.100]	[−0.061, 0.183]	[0.136, 0.174]	[0.083, 0.182]	[−0.006, 0.171]	[0.105, 0.226]	[−0.012, 0.141]

	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	0.149	0.025	0.090	0.117	0.101	0.148
	(0.053)	(0.070)	(0.083)	(0.049)	(0.061)	(0.056)
	[0.006]	[0.653]	[0.340]	[0.036]	[0.173]	[0.011]
t-test p-values	0.078		0.393		0.287	
R^2	0.194	0.247	0.319	0.140	0.298	0.241
Observations	292	194	133	351	228	242
File FE	✓	✓	✓	✓	✓	✓
Better Lee Bounds	—	−0.015, −0.011	—	0.096, 0.144	0.051, 0.113	0.103, 0.173
	—	[−0.110, 0.140]	—	[0.072, 0.171]	[0.016, 0.156]	[0.084, 0.192]

Notes: The dependent variable is a *Guarantor dummy* that equals "1" if the participant approves the credit application but requests a guarantor and "0"s if the participant approves it without requesting a guarantor. The sample is restricted to the first round of the experiment. When partitioning nonbinary 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 IAT. Higher IAT values indicate that participants associate men more with careers and women more with household tasks. The t-test p-value corresponds to one-sided tests. Romano-Wolf p-values are shown in square brackets and refer to Romano-Wolf stepdown p-values which control for the family-wise error rate and account for multiple hypothesis testing; we adjust for a pair of hypotheses in columns 1–2 and, separately, 12 hypotheses in columns 3–14 and follow Romano and Wolf's (2016) bootstrap re-sampling algorithm with 10,000 replications. *Better Lee Bounds* refer to Lee (2009) bounds that are tightened through a LASSO selection procedure that considers all participant covariates (Semenova 2021). Stoye (2009)–adjusted Imbens and Manski (2004) 95 percent confidence intervals are reported in brackets below these bounds. Cluster robust standard errors are shown in parentheses and clustered at the participant level. Table 1 contains all variable definitions.

forest reveals a broad distribution of treatment effects underlying the ATE. They vary from slightly negative to a 13 percentage points higher probability of requesting a guarantor when we present a loan application as coming from a female instead of a male entrepreneur.

Panel B of online Appendix Figure A5 ranks officer traits by their relative importance as moderators (drivers of treatment heterogeneity). We define a trait's relative importance as the weighted sum of the number of times it is used to split at each depth in the forest. The more a trait is used to split subsamples, the more predictive power it has. We find that loan officers' implicit bias against businesswomen, measured as their IAT score, is by far the main driver of treatment heterogeneity. In exactly a third of all trees, the algorithm picks an officer's implicit bias to make the first split. The second and third most important drivers are officer age and experience, which our algorithm—unlike linear regressions—can neatly disentangle. The other traits—risk aversion, gender, and hierarchical position—are much less important drivers of treatment heterogeneity. Most of these results are consistent with those based on split-sample regressions. Both show that implicit stereotypes, age, and experience are important, and they both tell us that loan officers' own gender is not an important driver of discriminatory guarantor requirements. An interesting exception is *Participant is supervisor*. Linear sample-split regressions suggest that this variable correlates strongly with bias in guarantor requirements. Yet, the causal forest tells us that this is not the case once we account for nonlinearities and the fact that being a supervisor correlates with age and work experience.

Figure 4 plots the predicted treatment effects against the three main officer traits. We fit smooth local polynomial functions in each scatterplot. The patterns are striking. Panel A shows how the predicted treatment effect increases when officers' implicit stereotypes are stronger. The causal forest reveals a discrete jump of 2.5 percentage points in the predicted treatment effect at an IAT score of 0.25. From a policy perspective, this indicates that there is a distinct group of biased loan officers that may be targeted by, for example, debiasing interventions. Panels B and C of Figure 4 show a tight negative correlation between age and work experience, respectively, and the predicted treatment effect. This relationship is much more linear. The probability that a loan officer engages in discriminatory guarantor requirements declines steadily with age and, independently, with work experience.

When two traits correlate strongly, an algorithm may arbitrarily pick one of them as a strong determinant of treatment heterogeneity while assigning a lesser role to the other. This can be problematic when interpreting the relative importance of moderators. Reassuringly, online Appendix Table A1 shows that most of our participant covariates are not highly correlated. As expected, the strongest correlations are between someone's age and their work experience and the probability of being a supervisor. The causal forest nevertheless selects both participant age and experience as two key drivers of heterogeneity. Even though these variables correlate, they contain sufficiently distinct information for the algorithm to prefer both of them to other variables (such as risk aversion and gender).[35]

IV. Interpretation and Mechanisms

In summary, when we present the same file as coming from a female entrepreneur instead of a male, officers are on average 6 percentage points (or 26 percent) more likely to require a guarantor. This biased behavior is concentrated among younger and less experienced loan officers and especially among those who harbor a stronger bias against female entrepreneurs. We now consider two mechanisms that may

[35] Mullainathan and Spiess (2017) discuss unstable feature selection in the context of traditional applications of LASSO and a "carefully constructed heterogeneity tree" (102). In contrast, our forest aggregates model fits from many thousands of trees. Each tree is fitted on a different random sample of observations, and the nodes in each tree consider a different random subset of variables. Because we average across a complete forest and track the weighted sum of the number of times a variable is used to split, we can more confidently discuss the relative importance of moderators.

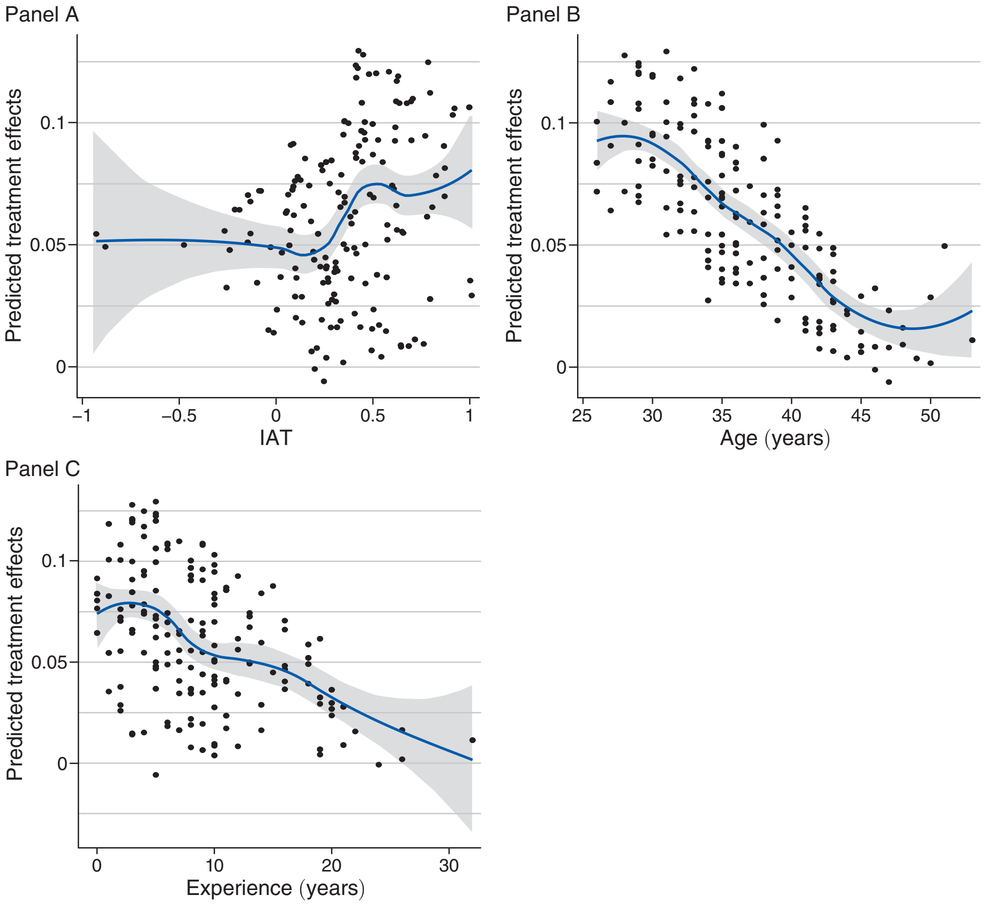

FIGURE 4. PREDICTED TREATMENT EFFECTS BY IMPLICIT GENDER BIAS, AGE, AND EXPERIENCE

Notes: Plotted points represent individual loan officers. The horizontal axis indicates implicit gender bias (IAT score, panel A), age (panel B), and experience (panel C). These are the three most important treatment moderators according to the causal forest algorithm (cf. online Appendix Figure A5, Panel B). The vertical axis in each panel indicates the CATE predicted by our causal forest. The lines display the local smoothed polynomial relationship between the loan officer trait and the CATE. The treatment effects are predicted by feeding our test sample (30 percent of the full sample) through the trees grown by the causal forest algorithm on the basis of the splitting sample (70 percent of the full sample).

underpin this result: gender differences in credit risk and loan officers acting on implicit biases that reflect social norms.

A. *Gender Differences in Credit Risk*

We first consider whether actual or perceived differences in credit risk could justify a different treatment of male and female applications. We offer several pieces of evidence that consistently show that the distribution of credit risk across male and female borrowers is very similar and, importantly, that loan officers themselves do not judge female borrowers to be riskier than equivalent male ones.

Gender Differences in Credit Scores.—We first compare the credit scores (from the credit registry) of the male and female applicants in our random sample of 250 loan applications. Recall that these were sampled from all applications that the bank received in recent years. The score captures an entrepreneur's borrowing and repayment history and is a good indicator of credit risk. The data reflect the real-life applications and the actual gender of the applicant, so they are nonexperimental. Since the sample is stratified by gender, firm size, region, and application quality, the distributions can be compared. The average score is 1,035 for men and 1,023 for women (a higher score implies less risk). This small difference is not statistically significant ($p = 0.80$). Online Appendix Table A4 presents OLS regressions for the 243 files for which credit scores are available (the dependent variable). The first column confirms that there is no significant difference between female and male applicants. This holds when we include sector fixed effects (column 2), add region fixed effects (column 3), and control for firm size (column 4) and amount requested (column 5). Online Appendix Figure A6 shows that the distribution of the credit scores is also very similar for male and female applicants (as confirmed by a Kolmogorov-Smirnov test).

Gender Differences in Subjective Repayment Probabilities.—Even if the distribution of ex ante credit risk is objectively very similar, loan officers may still perceive women to be riskier and hence be more demanding in terms of guarantor requirements. To see whether this is the case, Figure 5 shows a binned scatterplot of credit scores (horizontal axis) and loan officers' view of an applicant's repayment probability (vertical axis). Dark gray dots (light gray diamonds) show bin averages for loan applications presented as coming from male (female) entrepreneurs. Confidence intervals (95 percent) are based on a cubic regression spline of subjective repayment probability on the credit score.

Two messages emerge. First, we observe a tight correlation between credit score and subjective repayment probability along the risk distribution. When officers assess lower risk applications (higher credit scores), they systematically perceive these to have a higher repayment probability. Second, this tight correlation holds independently of whether we present a file as coming from a male or a female entrepreneur. This holds true along the risk distribution: at no point is there a statistically significant disconnect between how loan officers translate male versus female credit risk into subjective repayment probabilities. This is further corroborated by online Appendix Figure A7 and online Appendix Table A5. Figure A7 provides a Kernel density plot of the subjective repayment probability that loan officers assign to male and female versions of the same applications. Both distributions are very similar, as confirmed by a formal Kolmogorov-Smirnov test. Online Appendix Table A5 contains regressions similar to those in Tables 3 and 4 but with *Subjective repayment probability* as the dependent variable. As expected, there is no significant impact of the randomized gender of the loan applicant on the credit risk as perceived by loan officers themselves.

Gender and Risk: Evidence from a Separate Risk Module.—Next, we present evidence from a separate risk module that we implemented and in which officers

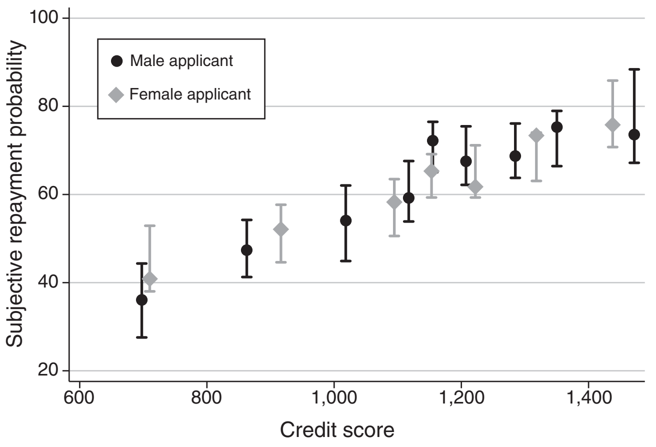

Figure 5. Credit Score and Subjective Repayment Probability, by Randomized Applicant Gender

Notes: This figure shows binned scatterplots for male applicants (dark gray dots) and female applicants (light gray diamonds) using robust pointwise confidence intervals. The data reflect all decisions in the first round of the experiment. The number of bins is not predetermined but data driven, and the integrated mean squared errors are minimized. The confidence intervals are at the 95 percent level and based on a cubic B-spline regression estimate of subjective repayment probability on the credit score. Credit scores are provided by the KKB credit registry, and higher scores indicate lower credit risk. Table 1 contains all variable definitions.

were randomly matched with a male or a female real-life entrepreneur. We informed the officers about the gender, age, and industrial sector of the person they had been matched with. Prior to the experimental sessions, we had asked these entrepreneurs to pick one out of six projects that were increasing in riskiness, in the spirit of Eckel and Grossman (2008). They had to do so for a project financed with debt and for one financed without debt. During the experiment, loan officers had to guess which risky projects their matched entrepreneur had chosen. We paid loan officers if they chose correctly.

The ordered probit specifications in online Appendix Table A6 regress the participants' perceptions of their matched entrepreneur's risk taking (on a 1–6 scale) on the gender of the entrepreneur. We control for the entrepreneur's age and industrial sector. For projects not funded with bank credit (column 1), loan officers believe that the entrepreneur they were matched with picked a slightly *less* risky project if that entrepreneur was female. The statistical significance of this gender difference disappears, however, when we ask loan officers about the risk they think entrepreneurs took for projects financed with bank credit (column 2). In either case, the evidence from this module is clearly at odds with officers perceiving female entrepreneurs to be *more* risky.

To sum up, we analyze objective credit scores, subjective repayment probabilities assigned by loan officers, and a module in which officers estimated the amount of risk taking by a real-life entrepreneur. Moreover, online Appendix B describes a second round of file reviews in which we experimentally varied the available applicant information. None of these exercises return compelling evidence supporting the hypothesis that gender differences in real or perceived credit risk can explain the strong gender bias in guarantor requirements that we document.

B. *Social Norms and Implicit Gender Bias*

Implicit Gender Bias and Guarantor Requirements across Industries.—We now investigate an alternative mechanism: implicit, norm-based biases that influence officers' decisions, especially when women apply for credit in gender-incongruent sectors. Decision-making can be biased when women are judged in stereotypically male domains.[36] We therefore investigate whether social norms and associated implicit biases present a credible mechanism to explain discriminatory guarantor requirements.

We first identify the two-digit ISIC industry of each of the 100 loan applications. This gives us 14 unique industrial sectors. We classify each sector as either a male-dominated or female-dominated one using data from the fifth and sixth rounds of the World Bank–EBRD Business Environment and Enterprise Performance Survey (BEEPS 2012–2020). This survey contains information on the gender of the owner of 44,540 firms across 48 middle-income countries in Emerging Europe, Central Asia, and North Africa.[37] For each industry, we measure the proportion of SMEs owned by women and then rank all industries. We define male-dominated (female-dominated) industries as those with a share of female-owned SMEs below (above) the median.[38] Examples of female-dominated sectors include the manufacturing of textiles and the manufacturing of food products and beverages, whereas male-dominated industries include the manufacturing of rubber and plastic products as well as construction.[39]

In the first two columns of Table 7, we test whether the substantially higher guarantor requirements for female loan applicants are equally present in male- and female-dominated industries. In case social norms play an important role, we would expect biased guarantor requirements to be mainly concentrated in male sectors. This is indeed what we find. In stereotypically male industries, the approval of a female loan application is almost 10 percentage points more likely to be made conditional on the presence of a guarantor (column 1). In stereotypically female industries, on the other hand, women entrepreneurs face no such bias (the coefficient is almost two times smaller and not statistically significant).[40]

[36] See, for example, Guiso et al. (2008); Carrell, Page, and West (2010); Reuben, Sapienza, and Zingales (2014); and Carlana (2019). Alan, Ertac, and Mumcu (2018) show how traditional gender views among Turkish elementary school teachers negatively affect girls' test performance.

[37] The survey uses a comprehensive sample frame (typically the business registry) of all formal private sector firms with at least five employees. The survey design ensures that the sample adequately represents the sectoral and geographical distribution of each country's SME population.

[38] Online Appendix Table A7 provides our sector breakdown and the male- versus female-dominated classification.

[39] In our sample, firms in female-dominated sectors are somewhat overrepresented. For example, the industry with the most files is ISIC 52 (retail trade, except of motor vehicles and motorcycles; repair of personal and household goods), which is a female-dominated industry and has 36 files. This is an artifact of stratifying by gender when we sampled the initial 250 files.

[40] When we randomize applicant gender, we create applications where the match between gender and industry is, by construction, artificial. Yet, the resulting applications reflect gender-industry combinations that are all observed in real life. Among the 250 files from which we draw our 100 loan applications, the percentage of male (female) applicants in male-dominated industries is 64 (36) percent. These numbers are 41 and 59 percent in female industries. This shows that while men (women) are clearly overrepresented in male-dominated (female-dominated) industries, there is sufficient overlap to create realistic experimental gender variation within both industry types. We also note that female applicants in male industries are not more or less risky—in terms of credit score—than male applicants in such industries (the p-value of a two-sided t-test for equal means is 0.90). The same holds for female industries (p-value $= 0.72$).

Table 7—Applicant Gender, Sectoral Gender Composition, and Guarantor Requirements

Dependent variable: Guarantor dummy	Male-dominated sectors	Female-dominated sectors	Male-dominated sectors		Female-dominated sectors	
			Below median IAT	Above median IAT	Below median IAT	Above median IAT
	(1)	(2)	(3)	(4)	(5)	(6)
Female applicant	0.098	0.054	−0.023	0.204	0.014	0.094
	(0.055)	(0.037)	(0.091)	(0.077)	(0.064)	(0.053)
t-test p-values	0.255		0.028		0.168	
R^2	0.114	0.166	0.248	0.352	0.306	0.274
Observations	219	564	108	106	277	271
File FE	✓	✓	✓	✓	✓	✓
Better Lee Bounds	0.081, 0.094	0.055, 0.057	−0.040, 0.029	0.161, 0.237	0.030, 0.082	0.030, 0.082
	[−0.030, 0.199]	[−0.023, 0.136]	[−0.157, 0.121]	[0.013, 0.379]	[−0.062, 0.176]	[−0.029, 0.135]

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. Female- and male-dominated sectors are defined by the share of firms with majority female ownership at the two-digit ISIC industry level using data from the EBRD–World Bank BEEPS V and VI. Female- (male-) dominated firms are those in industries with an above- (below-) median share of majority female-owned firms. *Better Lee Bounds* refer to Lee (2009) bounds that are tightened through a LASSO selection procedure that considers all participant covariates (Semenova 2021). Stoye (2009)–adjusted Imbens and Manski (2004) 95 percent confidence intervals are reported in brackets below these bounds. The sample is restricted to the first round of the experiment. Cluster robust standard errors are shown in parentheses and clustered at the participant level. Table 1 contains all variable definitions.

In columns 3 through 6, we split the decisions for stereotypically male sectors (columns 3–4) and for stereotypically female sectors (columns 5–6) into those taken by loan officers with a below-median IAT score (columns 3 and 5) and an above-median score (columns 4 and 6). For female-dominated sectors, we do not find a statistically significant difference between more and less implicitly gender-biased loan officers. In contrast, in male-dominated industries, the higher guarantor requirements for women are driven by loan officers with a strong implicit gender bias. Among these officers, there is a 20 percentage point gender difference in the probability of a guarantor request in stereotypically male industries. In unreported regressions, we find no relationship between applicant gender, on one hand, and subjective repayment probability in either male- or female-dominated industries on the other hand. This again indicates that the stricter guarantor requirements do not reflect officers’ concerns about higher credit risk for female applicants, even if these women apply in stereotypically male industries. Instead, our results offer strong support in favor of implicit biases, informed by social norms, underpinning our ATEs.

Implicit Bias, *Industries*, *and Guarantors: Heterogeneous Treatment Effects*.—We return to the causal forest to investigate heterogeneous treatment effects across industries. Online Appendix Figure A8 shows the distribution of the predicted treatment effects in female-dominated industries (dark gray bars) and male-dominated industries (light gray bars). We again observe a substantial spread in the conditional treatment effects around the ATEs. Interestingly, both distributions hardly overlap. Only the largest predicted treatment effects in female industries overlap with the smallest ones in male industries. This indicates that loan officers systematically judge female entrepreneurs differently—they apply a different standard—in male-versus female-dominated industries.

Online Appendix Figure A9 depicts the relative importance of officer traits as drivers of biased guarantor requirements in female-dominated industries (panel A) and male-dominated ones (panel B). The same traits as before play a key role: implicit bias (IAT score), age, and work experience. Figure 6 visualizes the stark difference between male and female industries in terms of the relationship between implicit gender bias (top panels), age (middle), and experience (bottom) and predicted treatment effects across loan officers. A first clear difference concerns implicit biases. In female sectors (left), individual treatment effects vary between −2.7 and 10.6 percentage points, but without an apparent relationship with officers' implicit bias. In contrast, in male-dominated sectors, the treatment effect is not only generally above 10 percentage points, but there is also a strong positive relationship between officers' implicit bias and their predicted discriminatory guarantor requirements. This illustrates how discrimination based on implicit biases about female entrepreneurs can be context dependent (Coffman 2014) and only manifests itself when women apply in stereotypically male sectors.

Strikingly, we observe the opposite pattern for loan officer age (middle) and work experience (bottom). The algorithm can disentangle the two and shows how both lead to a monotonic decline in biased lending behavior in female-dominated sectors. When officers reach an age of 45 or have about two decades of work experience, they typically no longer display a bias against female applicants—as long as these entrepreneurs stick to traditionally female industries.[41] In sharp contrast, the attenuating effect of age and experience is absent in male-dominated sectors (right). There, independent of an officer's age or experience, the predicted treatment effects consistently fluctuate between 10 and 15 percentage points.

V. Conclusions

We implement a lab-in-the-field experiment to gain insights into the nature of gender discrimination in small business lending. While we find no evidence of direct discrimination in terms of unconditional approval rates, we uncover a strong gender bias in loan requirements. All else equal, the approval of female applications is 26 percent more likely to be made conditional on the presence of a guarantor. A causal forest algorithm reveals that specific loan officer traits—their implicit bias about entrepreneurial women, their work experience, and their age—independently and strongly correlate with the intensity of discrimination.

What do these results tell us about the nature of the discrimination we observe? 'Classic' statistical discrimination does not appear to be a key mechanism. Several empirical exercises return no evidence that female and male applicants are objectively different or that loan officers hold different explicit beliefs about their riskiness. Instead, we show that officers with stronger implicit biases against women in

[41] Botelho et al. (2015) show how experience reduces racial discrimination by Brazilian teachers.

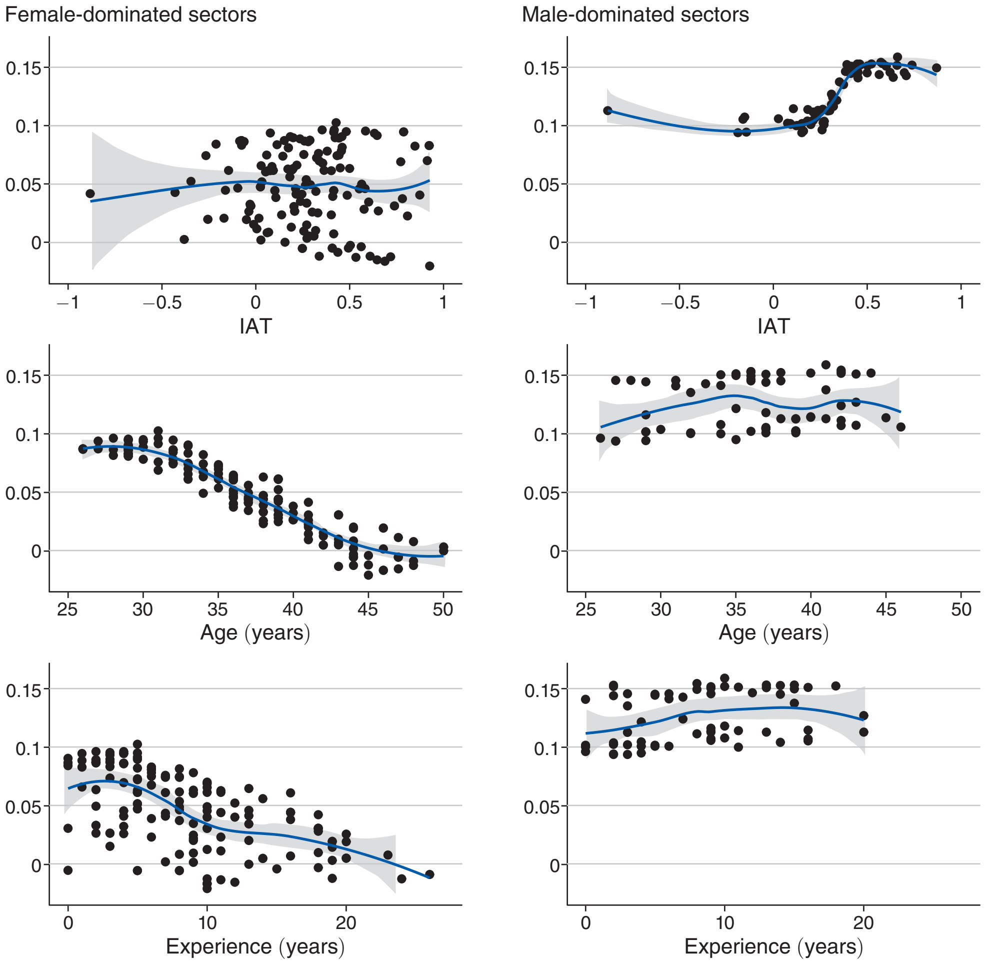

Figure 6. Predicted Treatment Effects across Sectors, by Implicit Bias, Age, and Experience

Notes: Plotted points represent individual loan officers. The horizontal axis indicates implicit gender bias (IAT score, top), age (middle), and experience (bottom). These are the three most important treatment moderators according to the causal forest algorithm (cf. online Appendix Figure A9). Female- and male-dominated sectors are defined by the share of firms with majority female ownership at the two-digit ISIC industry level using data from the EBRD–World Bank BEEPS V and VI. Female- (male-) dominated firms are those in industries with an above- (below-) median share of majority female-owned firms. The vertical axis in each panel indicates the CATE predicted by our causal forest. The lines display the local smoothed polynomial relationship between the loan officer trait and the CATE. The treatment effects are predicted for female- (male-) dominated sectors by feeding our test sample (30 percent of the sample corresponding to female- (male-) dominated sectors) through the trees grown by the causal forest algorithm on the basis of the splitting sample (70 percent of the sample corresponding to female- (male-) dominated sectors).

business make more discriminatory decisions in terms of guarantor requirements. While our empirical setup does not allow us to distinguish conclusively how implicit biases operate—directly, via statistical discrimination based on stereotypical beliefs, or via taste-based discrimination—we believe the latter channel is least likely. First, we would expect taste-based discrimination to already rear its head in the unconditional loan approval decisions, but it does not. Second, implicit bias mainly plays a role when women apply in gender-incongruent sectors—which is highly suggestive of a role of implicit biases steeped in social norms rather than reflecting individual animus.

Because biased guarantor requirements are concentrated among loans that perform well in real life, discrimination may be costly to the bank. If creditworthy female applicants cannot provide a guarantor, profitable projects go unfunded. In equilibrium, women may avoid applying for credit altogether—as they anticipate being asked for a personal guarantor. Moreover, in those cases where women can come up with a guarantor, there will be a cost for these entrepreneurs themselves as they are asked to put scarce social capital on the line. We provide survey evidence indicating that guarantor requirements are indeed regarded as a costly constraint by many Turkish businesswomen.

We sketch three courses of action for banks that want to mitigate gender discrimination among loan officers. First, our results show clearly that discrimination is less prevalent among older and more experienced loan officers (at least in female-dominated sectors). Adding more senior officers to relatively junior teams can then be a straightforward way to reduce the risk of discriminatory lending. Second, policies to mitigate the real-world impact of implicit biases may be called for. For example, banks can set branch-level goals for lending to women without a guarantor and hold those branches that do not meet this goal accountable. Successful female entrepreneurs can also be made more visible to loan officers—for instance, by integrating them in banks' internal communication and training programs. This holds in particular for female entrepreneurs in stereotypically male industries. Third, banks might consider replacing human decision-making with algorithmic decision-making altogether. But while algorithmic credit scoring can reduce face-to-face discrimination in markets prone to implicit and explicit biases, it may fail to reduce (or may even increase) disparities between and within social groups in lending terms (Bartlett et al. 2021; Fuster et al. 2021).

We end this paper with two observations about the generalizability of our findings. A first question is how well our lab results translate to the real-life Turkish setting. While loan officers were aware that their decisions were not "live" ones, we believe that the incentive scheme, combined with using real applications from the recent past, meant that day-to-day lending operations were simulated realistically in the lab. An interesting area for future research would be to mimic real life even more closely by integrating experimental elements into the regular lending decisions of loan officers.

A related question is how portable our results are across borders. One way to answer this is to identify countries that are similar to Turkey in terms of economic and financial development as well as gender norms.[42] This yields a broad and varied group of countries, including Egypt, Morocco, and the United Arab Emirates in the Middle East and North Africa; the Dominican Republic and Paraguay in Latin America; Greece and Hungary in Europe; and Cambodia and Sri Lanka in Asia.

[42] We identify the intersection of all countries within one standard deviation from Turkey in terms of GDP per capita, domestic credit to the private sector as a percentage of GDP, and the World Economic Forum Global Gender Gap Index.

In all these countries, discrimination by (parts of the) loan officer population may contribute to women's financial exclusion and, therefore, to a misallocation of entrepreneurial talent. Perhaps even more importantly, such discriminatory behavior will prevent banking systems from contributing to a fairer society with equal economic opportunities for all.

REFERENCES

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30 (2): 175–87.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–53.

Alan, Sule, Seda Ertac, and Ipek Mumcu. 2018. "Gender Stereotypes in the Classroom and Effects on Educational Outcomes." *Review of Economics and Statistics* 100 (5): 876–90.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics* 128 (2): 469–530.

Alesina, Alberto F., Francesca Lotti, and Paolo Emilio Mistrulli. 2013. "Do Women Pay More for Credit? Evidence from Italy." *Journal of the European Economic Association* 11 (Supplement 1): 45–66.

Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton, NJ: Princeton University Press.

Asiedu, Elizabeth, James A. Freeman, and Akwasi Nti-Addae. 2012. "Access to Credit by Small Businesses: How Relevant are Race, Ethnicity, and Gender?" *American Economic Review* 102 (3): 532–37.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *PNAS* 113 (27): 7353–60.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *Annals of Statistics* 47 (2): 1148–78.

Attanasio, Orazio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, and Heike Harmgart. 2015. "The Impacts of Microfinance: Evidence from Joint-Liability Lending in Mongolia." *American Economic Journal: Applied Economics* 7 (1): 90–122.

Baranov, Victoria, Ralph De Haas, and Pauline Grosjean. Forthcoming. "Male-Biased Sex Ratios and Masculinity Norms: Evidence from Australia's Colonial Past." *Journal of Economic Growth*.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. 2022. "Consumer-Lending Discrimination in the FinTech Era." *Journal of Financial Economics* 143 (1): 30–56.

Beck, Thorsten, Patrick Behr, and Andre Guettler. 2013. "Gender and Banking: Are Women Better Loan Officers?" *Review of Finance* 17 (4): 1279–1321.

Beck, Thorsten, Patrick Behr, and Andreas Madestam. 2018. "Sex and Credit: Do Gender Interactions Matter for Credit Market Outcomes?" *Journal of Banking and Finance* 87: 380–96.

Becker, Gary Stanley. 1957. *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.

BEEPS (Business Environment and Enterprise Performance Survey). 2012–2020. "EBRD–World Bank Business Environment and Enterprise Performance Survey." https://www.beeps-ebrd.com/data/ (accessed December 15, 2021).

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen. 2017. "Program Evaluation and Causal Inference with High-Dimensional Data." *Econometrica* 85 (1): 233–98.

Belloni, Alexandre, Victor Chernozhukov, Christian Hansen, and Damian Kozbur. 2016. "Inference in High-Dimensional Panel Models with an Application to Gun Control." *Journal of Business and Economic Statistics* 34 (4): 590–605.

Bellucci, Andrea, Alexander Borisov, and Alberto Zazzaro. 2010. "Does Gender Matter in Bank-Firm Relationships? Evidence from Small Business Lending." *Journal of Banking and Finance* 34 (12): 2968–84.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws. 2017. "Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment." *Journal of Finance* 72 (2): 509–38.

Bertrand, Marianne, and Esther Duflo. 2017. "Field Experiments on Discrimination." In *Handbook of Economic Field Experiments*, Vol. 1, edited by Esther Duflo and Abhijit Banerjee, 309–93. Amsterdam: Elsevier.

Bertrand, Marianne, and Sendhil 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.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. “Implicit Discrimination.” *American Economic Review* 95 (2): 94–98.

Blanchflower, David G., Phillip B. Levine, and David J. Zimmerman. 2003. “Discrimination in the Small-Business Credit Market.” *Review of Economics and Statistics* 85 (4): 930–43.

Blau, Francine D., and Lawrence M. Kahn. 2017. “The Gender Wage Gap: Extent, Trends, and Explanations.” *Journal of Economic Literature* 55 (3): 789–865.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. 2019. “The Dynamics of Discrimination: Theory and Evidence.” *American Economic Review* 109 (10): 3395–3436.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope. Forthcoming. “Inaccurate Statistical Discrimination: An Identification Problem.” *Review of Economics and Statistics*.

Bond, Philip, and Ashok S. Rai. 2008. “Cosigned versus Group Loans.” *Journal of Development Economics* 85 (1–2): 58–80.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. “Stereotypes.” *Quarterly Journal of Economics* 131 (4): 1753–94.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2019. “Beliefs about Gender.” *American Economic Review* 109 (3): 739–73.

Botelho, Fernando, Ricardo A. Madeira, and Marcos A. Rangel. 2015. “Racial Discrimination in Grading: Evidence from Brazil.” *American Economic Journal: Applied Economics* 7 (4): 37–52.

Brock, J. Michelle, and Ralph De Haas. 2023. “Replication data for: Discriminatory Lending: Evidence from Bankers in the Lab.” American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E157121V1.

Brooks, Alison Wood, Laura Huang, Sarah Wood Kearney, and Fiona E. Murray. 2014. “Investors Prefer Entrepreneurial Ventures Pitched by Attractive Men.” *PNAS* 111 (12): 4427–31.

Carlana, Michela. 2019. “Implicit Stereotypes: Evidence from Teachers’ Gender Bias.” *Quarterly Journal of Economics* 134 (3): 1163–1224.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. “Sex and Science: How Professor Gender Perpetuates the Gender Gap.” *Quarterly Journal of Economics* 125 (3): 1101–44.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernandez-Val. 2020. “Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India.” NBER Working Paper No. 24678.

Coffman, Katherine Baldiga. 2014. “Evidence on Self-Stereotyping and the Contribution of Ideas.” *Quarterly Journal of Economics* 129 (4): 1625–60.

Cole, Shawn, Martin Kanz, and Leora Klapper. 2015. “Incentivizing Calculated Risk-Taking: Evidence from an Experiment with Commercial Bank Loan Officers.” *Journal of Finance* 70 (2): 537–75.

Demirgüç-Kunt, Asli, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess. 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. Washington, DC: World Bank.

Demirgüç-Kunt, Asli, Patrick Honohan, and Thorsten Beck. 2008. *Finance for All? Policies and Pitfalls in Expanding Access*. Washington, DC: World Bank.

Diep-Nguyen, Ha, and Huong Dang. 2020. “Social Collateral.” Unpublished.

Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram Pathania. 2021. “Measuring Bias in Consumer Lending.” *Review of Economic Studies* 88 (6): 2799–2832.

Eckel, Catherine C., and Philip J. Grossman. 2008. “Differences in the Economic Decisions of Men and Women: Experimental Evidence.” In *Handbook of Experimental Economics Results*, Vol. 1, 509–19. Amsterdam: Elsevier.

Ewens, Michael, and Richard R. Townsend. 2020. “Are Early Stage Investors Biased against Women?” *Journal of Financial Economics* 135 (3): 653–77.

Field, Erica, Seema Jayachandran, and Rohini Pande. 2010. “Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India.” *American Economic Review: Papers and Proceedings* 100 (2): 125–29.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. “Predictably Unequal? The Effects of Machine Learning on Credit Markets.” *Journal of Finance* 77 (1): 5–47.

GADM. 2018. “Database of Global Administrative Areas, Version 3.6.” GADM. https://gadm.org/download_country.html (accessed February 24, 2022).

Glover, D., A. Pallais, and W. Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *Quarterly Journal of Economics* 132 (3): 1219–60.

Gompers, Paul A., and Sophie Q. Wang. 2017. "Diversity in Innovation." NBER Working Paper No. 23082.

Greenwald, A.G., D.E. McGhee, and J.L. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74 (6): 1464–80.

Greenwald, Anthony G., Brian A. Nosek, and Mahzarin R. Banaji. 2003. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality and Social Psychology* 85 (2): 197–216.

Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. 2009. "Understanding and Using the Implicit Association Test: III. Meta-analysis of Predictive Validity." *Journal of Personality and Social Psychology* 97 (1): 17–41.

Grosjean, Pauline, and Rose Khattar. 2019. "It's Raining Men! Hallelujah? The Long-Run Consequences of Male-Biased Sex Ratios." *Review of Economic Studies* 86 (2): 723–54.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. "Culture, Gender, and Math." *Science* 320 (5880): 1164–65.

Hebert, Camille. 2020. "Gender Stereotypes and Entrepreneur Financing." Unpublished.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and US Economic Growth." *Econometrica* 87 (5): 1439–74.

Imbens, Guido W., and Charles F. Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica* 72 (6): 1845–57.

Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F.P. Luttmer, and Kelly Shue. 2016. "Screening Peers Softly: Inferring the Quality of Small Borrowers." *Management Science* 62 (6): 1554–77.

Jaim, Jasmine. 2021. "Bank Loans Access for Women Business-Owners in Bangladesh: Obstacles and Dependence on Husbands." *Journal of Small Business Management* 59 (sup1): S16–S41.

Kaas, Leo, and Christian Manger. 2012. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment." *German Economic Review* 13 (1): 1–20.

Klapper, Leora, and Sandeep Singh. 2014. *The Gender Gap in the Use of Financial Services in Turkey*. Washington, DC: World Bank.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–1102.

McFarland, Sam G., and Zachary Crouch. 2002. "A Cognitive Skill Confound on the Implicit Association Test." *Social Cognition* 20 (6): 483–510.

Montoya, Ana Maria, Eric Parrado, Alex Solis, and Raimundo Undurraga. 2020. *Bad Taste: Gender Discrimination in Consumer Credit Markets*. Washington, DC: Inter-American Development Bank.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Naaraayanan, S. Lakshmi. 2020. "Women's Inheritance Rights and Entrepreneurship Gender Gap." Unpublished.

Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56 (3): 799–866.

OECD. 2016. *Entrepreneurship at a Glance 2016*. Paris, France: OECD.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4): 659–61.

Pozzolo, Alberto F. 2004. "The Role of Guarantees in Bank Lending." Unpublished.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How Stereotypes Impair Women's Careers in Science." *PNAS* 111 (12): 4403–08.

Romano, Joseph P., and Michael Wolf. 2005. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73 (4): 1237–82.

Romano, Joseph P., and Michael Wolf. 2016. "Efficient Computation of Adjusted p-values for Resampling-Based Stepdown Multiple Testing." *Statistics and Probability Letters* 113 (38): 38–40.

Safavian, Mehnaz, and Aban Haq. 2013. *Are Pakistan's Women Entrepreneurs Being Served by the Microfinance Sector?* Washington DC: World Bank Group.

Semenova, Vira. 2020. "Generalized Lee Bounds." *arXiv*: 2008.12720.

Singh, Sanjana, Radhika Asrani, and Anupama Ramaswamy. 2016. *Study on Mapping the Market Potential and Accelerating Finance for Women Entrepreneurs in Bangladesh*. Washington, DC: International Finance Corporation.

Stiglitz, Joseph E. 1990. "Peer Monitoring and Credit Markets." *World Bank Economic Review* 4 (3): 351–66.

Stoye, Jörg. 2009. "More on Confidence Intervals for Partially Identified Parameters." *Econometrica* 77 (4): 1299–1315.

Tibshirani, Julie, Susan Athey, Rina Friedberg, Vitor Hadad, David Hirshberg, Luke Miner, Erik Sverdrup, and et al. 2020. "GRF Package." https://grf-labs.github.io/grf/.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Efects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42.

World Economic Forum. 2018. *The Global Gender Gap Report*. Geneva, Switzerland: World Economic Forum.