Discrimination in the Venture Capital Industry: Evidence from Two Randomized Controlled Trials*

Ye Zhang †

December 27, 2021

Abstract

This paper examines discrimination based on startup founders' gender, race, and age by early-stage investors using two complementary field experiments with real U.S. venture capitalists. The first experiment invites investors to evaluate multiple randomly generated startup profiles, which they know to be hypothetical, in order to be matched with real, high-quality startups from collaborating incubators. Investors can also donate money to randomly displayed startup teams to show their anonymous support during the COVID-19 pandemic. The second experiment consists of sending hypothetical pitch emails with randomized startups' information to global venture capitalists and comparing their email behaviors by utilizing a new email technology that tracks investors' detailed information acquisition behaviors. These experiments provide the following findings. (i) In the contact stage, investors are biased towards female, Asian, and older founders of relatively low-quality startups, while they are biased against female, Asian, and older founders of relatively high-quality startups. This result provides a potential reconciliation of the contradictory results in the extant literature. (ii) Among multiple coexisting sources of bias identified, statistical discrimination and implicit bias are important reasons for investors' "anti-minority" behaviors. A consistent estimator is developed to measure the polarization of investors' biases and their separate driving forces. And finally, (iii) there was a temporary, stronger bias against Asian founders during the COVID-19 outbreak, which started to fade in April 2020. *JEL Classification:* C93, D83, G24, G40, J15, J16, J71.

*I would like to express my deepest appreciation to Jack Willis, Harrison Hong, and Wei Jiang for their guidance, support, and profound belief in the value of my work. I am also grateful to Donald Green, Sandra Black, Mark Dean, Alessandra Casella, Matthieu Gomez, Jose Scheinkman, Eric Verhoogen, Jushan Bai, Junlong Feng, Michael Best, Bentley MacLeod, Alexander Ljungqvist, Ulrike Malmendier, Bernard Salenie, Per Strömberg, Corinne Low, Shi Gu, Andrew Prat, Xavier Giroud, Johannes Stroebel, Olivier Toubia, and Patrick Bolton for their valuable comments. I thank the participants at the PhD Colloquiums at Columbia University and the investors who participated in these experiments. Special thanks go to Corinne Low, Colin Sullivan, and Judd Kessler for sharing their IRR Qualtrics code package. This project was supported by PER funding from the Columbia University Economics Department, by the Columbia University Eugene Lung Entrepreneurship Center Fellowship, and by the Columbia CELSS Seed Grant. The project is registered at AEA RCT Registry (AEARCTR-0004982). All errors are my own. Total word count: 16,721.

†Stockholm School of Economics Finance Department. Email: Ye.Zhang@hhs.se

1 Introduction

The persistent gender gap and racial disparities (Bertrand, Goldin and Katz (2010), Bertrand and Hallock (2001), Chetty, Hendren, Jones and Porter (2020)) at the top of the earnings distribution often raise the question of whether explicit or implicit discrimination creates a glass ceiling for women and nonwhites in modern U.S. society (Bertrand, Chugh and Mullainathan (2005a)).[1] Therefore, a significant amount of attention and debate about discrimination concentrate on top-level, high-skilled labor markets and relevant financial markets.[2] Despite the existence of a large body of academic literature on empirically testing discrimination in various markets,[3] most prior work mainly focuses on the entry-level and low-skilled labor market and product market (Bertrand and Mullainathan (2004), Giuliano, Levine and Leonard (2009), Kessler, Low and Sullivan (2019), List (2004)). In this paper, I focus on the U.S. entrepreneurial finance market, which generates huge amounts of wealth and many business leaders. Through two field experiments with real U.S. venture capitalists, this paper investigates whether early-stage investors are biased against female founders, Asian founders, and older founders when selecting startups for investment.[4] Since the VC industry plays an important role in fostering innovative and successful companies (Bernstein, Giroud and Townsend (2016)), identifying such discrimination is of critical importance not only for maintaining social fairness (Fang and Moro (2011)) and assessing the efficiency (Bertrand (2020)) of capital allocation in high-impact startups, but also for explaining the persistent gender and racial gap at the top level.

While the stark gender funding gap and the less favorable treatment received by nonwhite founders in the fundraising process has been well-documented (Ewens and Townsend (2020), Guzman and Kacperczyk (2019), Henderson, Herring, Horton and Thomas (2015), Hebert (2020)),[5] identifying discrimination and its nature in this setting poses new challenges to the commonly used empirical methods.[6] Furthermore, extant literature provides seemingly conflicting results. On the one hand, papers exploiting regression-based methods (Ewens and Townsend (2020), Guzman and

[1] Hegde, Ljungqvist and Raj (2021) document the "glass ceiling" phenomenon by analyzing detailed administrative records for the universe of patent examiners at the U.S. Patent and Trademark Office. They find that minority examiners are less likely to be promoted to the most senior grades.

[2] "Having A Glass Ceiling To Break Through Is Privilege. Here's Why." May 13, 2021, Forbes

[3] For a recent summary, see Lang and Kahn-Lang Spitzer (2020), Lang and Lehmann (2012), and Bertrand and Duflo (2017).

[4] Based on Gompers and Wang (2017a), 87% of U.S. venture capitalists are white, and investors may also have an unconscious bias against minority founders. Given the uniqueness of the entrepreneurial financing setting, this paper mainly studies racial bias about Asians, who are the largest minority group in the U.S. entrepreneurial community. "Asians" in this paper primarily stands for "East Asian" groups who originate mainly from China, Korea, Vietnam, etc. According to Gompers and Wang (2017a), Asians account for 18% of new U.S. venture capitalists and 15% of new entrepreneurs entering the market. Studying discrimination against African Americans and other under-represented minorities is an important question. However, my experimental design needs to be adjusted for future researchers to study these important questions.

[5] Gompers and Wang (2017b) demonstrate that from 1990-2016, women have made up less than 10% of the entrepreneurial and venture capital labor pool, which has the contrasting pattern of an increase in female labor market participation. Based on Gornall and Strebulaev (2020a), venture capitalists only invested 1 dollar in startups with female founding teams for every 35 dollars invested in startups with male founding teams in 2017. Also, Guzman and Kacperczyk (2019) document that female-led ventures are 63 percent less likely than male-led ventures to have obtained external funding (i.e., venture capital) from 1995-2001 even though women and men are equally likely to achieve exit outcomes through IPOs or acquisitions.

[6] According to List (2004), empirically testing for marketplace discrimination has taken two quite distinct paths in economic research: regression-based methods and field experiments.

Kacperczyk (2019), and Henderson et al. (2015)) often show that early-stage investors are biased *against* female and Asian founders. These non-experimental studies can suffer from omitted variable bias problems if they do not exploit exogenous variations.[7] However, a natural experimental setting is rarely available to solve various endogeneity concerns. On the other hand, Gornall and Strebulaev (2020a) implement the first correspondence test, a widely used field experiment method, in the U.S. VC industry. They find that early-stage investors are biased *towards* female and Asian founders. By sending out fictitious pitch emails with randomized email senders' names to U.S. venture capitalists, they compare investors' email response rates and find that, surprisingly, investors reply more frequently to emails sent by female and Asian names. Applying the correspondence test to the entrepreneurial finance market brings new challenges.[8] First, sending cold call emails is not the mainstream fundraising method used by high-quality startups. Hence, results may mainly capture how investors evaluate low-quality startup teams. Second, the "low-response-rate" problem is more severe in this setting compared with other markets (Bartoš, Bauer, Chytilová and Matějka (2016)), making it difficult for researchers to introduce variations in startup quality and test the nature of discrimination.[9]

To overcome the challenges mentioned above and reconcile the mixed results in the literature, this study implements two complementary field experiments in the U.S. entrepreneurial finance market, referred to as Experiment A and Experiment B in this paper. Experiment A is a relatively high-stakes experiment which uses real investment opportunities to recover investors' revealed preferences and directly identify the nature of discrimination.[10] To implement Experiment A, I collaborate with several accelerators and build a "Nano-Search Financing Tool," a machine learning matching tool composed of the following two parts. In the first part of this matching tool, to test any belief-driven bias, I invite real U.S. investors to evaluate multiple randomly generated startup profiles. Investors know the profiles are hypothetical, but they are willing to provide truthful evaluations so that the algorithm is better able to help them find real matched investment opportunities. Some randomly selected investors also receive a "monetary incentive" following Armona, Fuster and Zafar (2019) so that the more accurate investors' evaluation results are, the larger the monetary award the lottery winners will receive. This part essentially follows the incentivized resume rating (IRR) experimental paradigm created by Kessler et al. (2019). In the second part of this matching tool, to test any taste-driven bias, the tool provides each investor with an unexpected $15 Amazon Gift Card. Investors can accept it or anonymously donate a portion of the $15 to randomly displayed startup teams. Investors are also told that the researcher will use

[7] Since most venture capital funds and startup companies are private firms, startups' unique comparative advantages are usually only observable to investors rather than to researchers. Also, most databases only observe the match outcomes between investors and startups, making it difficult to separate investors' preferences from founders' preferences.

[8] Standard limitations of the correspondence test are discussed in Bertrand and Duflo (2017).

[9] According to Gornall and Strebulaev (2020a), the response rate to cold call pitch emails is about 6.5% in 2018 even though all the emails were designed to be as attractive as possible. This paper shows that during the COVID-19 crisis, the email response rate is only about 1%.

[10] Disentangling the nature of discrimination requires researchers to separate various belief-based reasons (i.e., "statistical discrimination") (Bertrand and Duflo (2017), Altonji and Blank (1999), Phelps (1972), Arrow et al. (1973)) from different taste-based reasons (i.e., "animus") (Becker (2010)). This disentanglement is difficult in the discrimination literature (Gneezy, List and Price (2012)) despite its importance in both welfare analysis and policy-making (Bohren, Imas and Rosenberg (2019a), Neumark (2012)).

the donated money to purchase small gifts for the corresponding real startup teams in the collaborative incubators to give founders some encouragement during the COVID-19 pandemic. This part essentially follows the dictator game experimental design (Carpenter, Connolly and Myers (2008)), which is widely used in lab experiments.

Experiment A's results show the existence of investors' bias and reconcile the contradictory results in the literature with the following findings: i) this experiment does not discover group-level *explicit* bias against minority founders when analyzing all the evaluated profiles. However, it identifies *implicit* bias against female and Asian founders whose magnitude is more than 40% of the effect of going to an Ivy League college in investors' evaluations.[11] When investors are fatigued after evaluating multiple profiles, their investment interest in female founders and the quality evaluations of both female and Asian founders start to decline significantly. Specifically, tech investors consider female founders' startups to be less profitable, consistent with "statistical discrimination". ii) the direction of implicit discrimination can reverse, depending on context. This phenomenon provides a new explanation to reconcile contradictory results in existing literature.[12] In the contact stage, investors are biased *against* minority founders of "high-quality" startups while biased *towards* minority founders of "low-quality" startups;[13] iii) taste-driven homophily exists as male (female) investors donate more to male (female) founders.[14] With the maximum donation equal to $15, male investors on average donate $3 less to female founders compared to similar male founders.[15]

Experiment A shows that some investors have implicit biases against minority founders. However, there are also some impact funds that support minority founders in the sample. So, how divided is the investment community in terms of their attitude towards minority founders, and what separates us? To answer this question in Experiment A, I develop a consistent "decision-based heterogeneous effect" estimator by using the "leave-one-out" technique and exploiting the exogeneous "within-individual" level randomization.[16] Standard heterogeneous effects rely on participants' predetermined demographic information. However, since each investor evaluates multiple randomized profiles, researchers can identify "individual-level" preferences and classify the participants into "anti-minority" groups and "pro-minority groups" based on their indicated decisions. The estimator finds that the split between investors' attitudes towards female founders is larger than those towards Asian and older founders. For gender bias, investors who prefer not contacting female founders expect that women-led startups have 16.40 percentile ranks lower potential financial returns

[11] According to Bertrand, Chugh and Mullainathan (2005b), implicit bias refers to the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner. In Experiment A, the investment interest in female founders and the quality evaluations of both female and Asian founders significantly decline when investors are fatigued.

[12] Another possible explanation to reconcile the literature is that investors do not exhibit implicit bias when replying to emails, which is also consistent with the results of Gornall and Strebulaev (2020a). However, Experiment B finds that investors check emails at different times during the day, which makes this explanation less plausible.

[13] The quality of a startup is measured by how likely the investor is to contact it.

[14] "Homophily effect" refers to the tendency for people to seek out or be attracted to those who are similar to themselves.

[15] Consistent with DellaVigna, List, Malmendier and Rao (2013), I also find that male investors, on average, are more generous than female investors.

[16] Junlong Feng provides crucial help and discussions for developing this estimator.

than men-led startups. However, investors who prefer contacting female founders expect that women-led startups have 7.93 percentile ranks higher potential financial returns than men-led startups. Therefore, holding different beliefs is an important reason for this split in attitudes towards female founders. Similarly, the decisions of the "anti-Asian" and "anti-older" groups, who prefer not contacting these startup founders, are also mainly affected by their beliefs that these startups are not profitable.

To check the external validity of Experiment A and determine why investors are biased towards minority founders of "low-quality" startups, I redesigned the correspondence test, referred to as Experiment B in this paper. This experiment also enables researchers to identify the nature of discrimination when the response rate is low.[17] During the COVID-19 outbreak (03-04/2020), I sent hypothetical pitch emails to more than 17,000 global venture capitalists with randomized founder names indicative of gender and race, randomized founder educational backgrounds, and randomized startup project characteristics displayed in both the emails' *subject lines* and in the emails' contents. By utilizing new email tracking technology, I can monitor detailed information acquisition behavior for each investor, including email opening behavior, time spent on pitch emails, click rate on the inserted startup's website, the contents in email replies and the email response rate. This experimental design and behavioral measurements generated enough experimental power to survive in the harsh experimental environment of the pandemic, when early-stage investors dramatically slowed down their investment pace (Howell, Lerner, Nanda and Townsend (2020)).

Experiment B's results confirm that investors are slightly biased *towards* female and Asian founders in the pitch email setting, which possibly captures investors' preferences when evaluating less attractive startups. Consistent with Gornall and Strebulaev (2020a), using minority names generally increases the email opening rate by roughly 1% compared to using majority names. After further testing the nature of bias, I find that the bias towards female founders is likely driven by taste-based reasons as the "female name" effect is much larger for impact funds. However, the bias towards Asians is likely driven by belief-based reasons because revealing more information about the team's quality can shrink this bias.[18] As Experiment B was implemented during 2021/03-2021/04, I find a temporary, strong bias against Asian founders during the COVID-19 outbreak. Investors spent 24% less time on pitch emails sent by Asian names compared to white names in March 2020. However, this bias quickly reversed starting in April of 2020. Hence, people's biases can also be temporarily affected by big societal events.

The contribution of this paper is both empirical and methodological. Empirically, this study provides experimental evidence that confirms the existence of early-stage investors' implicit bias against minority founders and tests its

[17] External validity refers to how generalizable the findings are in other settings, among a larger population, or at different times.

[18] After indicating that the founder has graduated from Ivy League colleges, I find investors' bias towards Asians starts to decline. This provides suggestive evidence of "belief-based" mechanisms.

underlying sources. Furthermore, it also documents a temporary, stronger bias against Asian founders during the COVID-19 outbreak, showing how discrimination can be affected by big societal events. Lastly, this paper provides suggestive evidence that reconciles contradictory results in the literature by showing how the direction of bias depends on context. Therefore, this paper empirically contributes to both discrimination literature and entrepreneurial financing literature.

Methodologically, the main contribution of this study is to provide a framework to identify discrimination and its nature in the financial market after solving the challenges faced by commonly used methods through a series of redesigned field experiments. Moreover, the developed decision-based heterogeneous effect estimator using "within-individual" level randomization helps to measure how divided society is and what separates us in terms of people's attitudes towards minorities.

To the best of my knowledge, this is the first paper to implement the correspondence test experiment and the IRR experiment together and compare their results. The IRR experimental paradigm, which is an incentivized elicitation technique invented by Kessler et al. (2019),[19] is motivated by providing a more ethical experimental design that can substitute for the standard correspondence test involving deception. By comparing the results from these two experimental methods, this paper demonstrates the validity of the IRR experimental method and its ability to identify subtle mechanisms, test heterogeneous effects and distributional effects, and generate results about later-stage decisions.[20] Notwithstanding these impressive merits, the IRR experiment is likely to be a good complementary experimental design rather than a full substitute for audit studies or correspondence tests due to the sample selection bias during the recruitment process and the potential consent form effect. I leave addressing these limitations to future research.

This paper is organized as follows. Section 2 discusses the construction of the individual-level VC investor database. Section 3 presents the design of Experiment A and analyzes investors' evaluations of startup profiles. The distributional effect discovered helps to reconcile the mixed results in the literature, and the decision-based heterogeneous effect estimator developed in this experimental setting measures the polarization of investors' biases and their separate driving forces. Section 4 describes the design of Experiment B and analyzes investors' information acquisition behaviors. Section 5 discusses the complementarity of these two experiments and the related policy implications. Section 6 concludes.

[19] Thanks to Corinne Low for insightful discussions clarifying the following important nature of the IRR experiment. Following the widely accepted Becker-Degroot-Marschak elicitation techniques of willingness to pay, the IRR experiment provides an incentive structure for eliciting true preferences and provides within-individual level exogenous variation. Also, the primary context of the IRR experiment is usually non-experimental, and subjects' motivation for participating in the study is mainly to receive the commercial benefits. Unlike a "survey," IRR experiment implementation requires much more social resources in order to reveal true preferences and generate causal evidence.

[20] Correspondence test generally reveals whether evaluator want to contact candidates. However, IRR can also reveal their investment preferences.

2 Data

To implement these field experiments, I have constructed a cross-sectional, individual-level global venture capitalist database, which contains the updated demographic information and contact information for 17,882 investors before 02/2020.[21] Since all experiments are implemented in English, only investors from English-speaking areas are included. This database combines multiple commercial databases and manually collected data. Detailed database descriptions and the key variable construction process are provided in Online Appendix A.

Despite the granular information provided by this database, it is important to realize the following three limitations. First, this database contains systematically more investors from the U.S. as well as more senior VCs due to data availability online and the data collection method used by data companies.[22] Hence, it may not be representative of the true geographical distribution of all venture capitalists in the world. Second, because of the high turnover rate within the VC industry, the contact information and status of these investors need to be updated frequently before use. Third, except for the key variables like gender, seniority, and location, other demographic variables are only available for relatively famous investors whose biographies are more readily available online.

The Summary Statistics of the 17,882 investors' demographic information is provided in Table 1. Panel A reports the location distribution of these investors, showing that U.S.-based investors account for 84.91% of this set of investors.[23] Panel B shows that most investors are interested in the Information Technology industry. Other important preferred industries include Healthcare, Consumers, and Energy. Panel C summarizes investors' background information. On average, female investors account for 24% of total investors. This is consistent with the NVCA/Deloitte survey results showing that women accounted for 21% of investment professionals in the U.S. VC industry in 2018 due to recent progress in increasing diversity.[24] Senior investors, who are partners, president, C-level managers, or vice president and above, account for 84% of total investors in our sample. Most investors are institutional investors, and angel investors only account for 11% of our sample. Moreover, only 2% of all investors work in not-for-profit impact funds.[25] If I use indicative keywords in the fund descriptions to classify the VC funds following Barber, Morse and Yasuda (2020), this percentage increases to 6%-8% depending on the keyword selection method.

[21] All investors' email addresses had been verified by a testing email before the experiments started.

[22] Most of the commercial databases used here are provided by U.S. data companies and collected by English speakers except for Zdatabase, which is the most comprehensive and timely database covering VC and PE activities in China.

[23] Maps of investors' global geographical distribution and U.S. geographical distribution are provided in Online Appendix Figure A1 and A2.

[24] See https://www2.deloitte.com/content/dam/Deloitte/us/Documents/audit/us-audit-egc-nvca-human-capital-suvey-2018.pdf. Gompers, Mukharlyamov, Weisburst and Xuan (2014) also show that women are under-represented among senior investment professionals in the VC industry.

[25] Pitchbook classifies VC funds into not-for-profit funds and for-profit funds together with the description of their investment preferences.

3 Experiment A's Design and Results

The goal of designing Experiment A as presented here is to elicit investors' investment preferences for startups of various qualities with a stronger incentive (i.e., real investment opportunities). To further directly test the nature of any detected biases, Experiment A combines the following two preference elicitation techniques: i) the IRR experiment, designed to directly test belief-based discrimination mechanisms, and ii) the dictator game, designed to directly test taste-based discrimination mechanisms. This section is organized as follows. Section 3.1 introduces the experiment's design and implementation details. Section 3.2 analyzes investors' evaluation results and donation decisions. Section 3.3 discusses the experiment's limitations.

3.1 Experimental Design

A. Investors and Recruitment

Experiment A was implemented from 03/2020 - 09/2020 using only online recruitment methods. I sent invitation emails together with instruction posters to the 15,000+ U.S. venture capitalists who also participated in Experiment B during the same period.[26] This ensures comparability across these experiments. To recruit real venture capitalists and create an experimental setting that closely mimics the real world, I have partnered with several real incubators and we have built a machine learning, algorithm-based matching tool, called "Nano-Search Financing Tool". Developing these data-driven matching tools has become popular in the VC industry. Incubators and VC funds, such as Techstars, Social+ Capital, and Citylight Capital, have done extensive work on developing machine learning algorithms to help evaluate investments, seek deals, and complement existing mainstream due diligence methods.[27] Similarly, our tool aims to match investors with startups in the collaborating incubators based on investors' revealed preferences. The dictator game in the donation section also provides important insights as venture capitalists add value to startups not only by offering funding, but also by providing their own advice and other types of support.

During the recruitment process, both the recruitment emails and posters emphasize that investors are invited to try a machine learning "Startup-Investor Matching Tool" that helps to identify startups that match with investors' pref-

[26] During the pandemic, Columbia IRB paused all field work which involves person-to-person activities due to COVID-19. The recruitment email templates and the instruction poster templates are provided in Online Appendix B Figure B9, Figure B10, Figure B11 and Figure B12

[27] Several important startup characteristics, such as the founder's passion and confidence, cannot be fully quantified by data. Hence, these data-driven methods are usually designed to complement person-to-person multiple stage investment strategies and are generally used for the initial screening process.

erences. Moreover, most investors also indicate that they are attracted by the potential deals, given that it takes more than 40 minutes to finish the experiment. Although we also inform investors of the related research purposes as required by IRB, this is framed as secondary. Therefore, this study has the ecological validity of a "natural field experiment," except that the subjects also know that their anonymized data will be analyzed by researchers.

There are, in total, 69 real U.S. investors from 68 different funds participating in this project, which provides 1,216 startup profile evaluation results.[28] The number of recruited experimental participants is comparable with Kessler et al. (2019). One important advantage of the IRR experimental design lies in the fact that researchers can obtain a large enough sample size despite recruiting a relatively small number of participants. This is a significant advantage in environments where it is hard to recruit many subjects, such as in the financial market where practitioners are extremely busy.

Similar to many other field experiments, Experiment A, with roughly a 0.5% response rate, suffers from sample selection bias during the recruitment process. Table 2 reports the summary statistics of participants' observable background information, showing that our sample investors are more likely to come from larger VC funds and to be minority founders. The average assets under management (AUM) of the VC funds in my sample is $547.46 million, which is larger than the average AUM of $444.44 million in the U.S. VC industry in 2019 based on an NVCA survey.[29] 42% of investors in the sample belong to minority groups (i.e., Asian, Hispanic, African, etc.), which is higher than the percentage of minority investors in the U.S. VC industry. However, the sample investors are representative in many other dimensions. For instance, recruited investors are mainly early-stage investors with preferences covering almost all major industries that VCs focus on. 86% of recruited investors are in senior positions, and about 20% are female. This is consistent with the situation described by the global investors' database.

Sample selection bias can also arise for the following unobservable reasons. First, participants can be more pro-social and willing to help research studies. Second, they may have a preference for Ivy League universities because the research project discussed was supervised by Columbia University, a member of the Ivy League. Third, recruited investors may be interested in understanding how data-driven methods can help investment evaluations. Such sample selection bias does not hurt the experiment's internal validity, yet it implies that it is important to implement complementary experiments in different settings in order to check the external validity.

[28] Recruiting *real* venture capitalists is crucial to understand professionals' startup investing strategies because the valuation of startups requires very specific skills(Gornall and Strebulaev (2020b)). Floyd and List (2016) and Carpenter et al. (2008) also emphasize the importance of recruiting professionals in finance-related experiments.

[29] During the recruitment period (i.e., the COVID-19 recession), only larger funds still have the money to look for new investment opportunities, whereas most smaller VC funds have shifted to "survival mode" where they focus on helping the startups they are currently investing in to survive rather than "purchasing" new undervalued startups in 2020. Based on many investors' email replies, shifting to "survival mode" is the main reason that they cannot participate in this research.

B. Survey Tool Structure

If investors are interested in participating in this experiment, they need to open the link inserted in the recruitment email to start the online Qualtrics survey using their browsers. The survey tool contains the following two sections. After reading the consent form, investors first enter the profile evaluation section (i.e., the IRR experiment section), where they need to evaluate 16 randomly generated startup profiles and answer standard background questions.[30] The second section is the donation section (i.e., the dictator experiment section), where investors decide how much of an unexpected $15 Amazon Gift Card they want to donate to randomly displayed startup teams. Figure 1 provides a flowchart for the experiment to demonstrate the tool's structure.

To help participants understand how the incentive works, I also provide an instruction page before the first profile evaluation section. Although investors know that all the startup profiles are hypothetical, this instruction page emphasizes that "the more accurately they reveal their preferences, the better outcomes the matching algorithm will generate (and the higher financial return that the lottery winner will obtain)." Moreover, since most VC investors only invest in startups in their industries and stages of interest (called "the quality/disqualify" test), I ask all the participants to assume that the generated startups they will be evaluating are in their industries and stages of interest.[31]

B.1 Profile Evaluation Section (IRR Experiment)

Following the factorial experimental design, I randomize the startup founder's gender, race, age, and multiple other startup characteristics simultaneously and independently in each created startup profile. I first create a set of team characteristics (such as the founding team's educational background), project characteristics (such as market traction) and existing financing situations. Then the backend Javascript code will randomly draw different characteristics and combine them together to create a hypothetical startup when each participant evaluates a new startup profile.[32]

In order to make the profiles look more realistic, I implement the following two designs. First, each hypothetical startup

[30] At the beginning of the study, each investor evaluated 32 profiles, and 6 investors finished the 32-profile version of the evaluation task. However, to recruit more investors, later participants only need to evaluate 16 profiles. Results are similar after removing the first 6 investors.

[31] Generally, investors need to go through the startup's industry, stage, and prior market knowledge to tell whether the startup is worth looking at. Another potential way to pass the "quality/disqualify" test is to provide several survey questions asking about investors' preferred industry, stages, and even revenue range before the evaluation section as Kessler et al. (2019) did. I did not do this due to the following two reasons. First, the market changes very quickly in the entrepreneurial community. Hence, it is hard to predict whether the startup information created in the design stage is still valid when I send out the invitation emails. Such a situation happened during the COVID-19 period when multiple industries got hit badly within a short period. Second, from the research perspective, I need insights from investors focusing on different areas and industries. This requires that the information provided should be general enough to accommodate as many participants with diverse backgrounds as possible.

[32] Sometimes the random combination may generate unusual cases like a startup with 50+ employees still not generating profits (see Amazon's history). However, such cases account for a small percentage of total generated cases.

profile is constructed using real startup components with ranges based on data from Pitchbook Database. Second, the information provided follows a "Crunchbase format" and is usually publicly available on LinkedIn, Crunchbase, or AngelList.[33] Some investors like Plug and Play Tech Centers sometimes go to these public platforms to look for relevant startups that fit their portfolios. The current design is meant to mimic this type of startup seeking behavior and provides data-driven methods for pre-selection decisions.[34] To test implicit bias, I also deliberately introduce a short break after investors evaluate the first half of the startup profiles (i.e., the first 8 profiles). This break provides a page indicating the experimental progress and encouraging the investor to finish all the evaluations. All the randomization of different startup components is provided in Table 3. The detailed construction process of the startup characteristics is provided in B.

Manipulating Gender and Race. — To indicate the gender and race of the startup founder, I randomly assign each hypothetical startup team member a first name highly indicative of gender (male or female) and a last name highly indicative of race (Asian or white). To make such information more salient, all the members in the same startup team are assigned names of the same gender and race. Moreover, I also include the founder's name in the evaluation questions and use indicative words like "she/her/his/him/he". Similar to other components, the combination of first names and last names is dynamically implemented by Qualtrics.[35] The detailed name selection process and the list of full names used in the tool are provided in Online Appendix B Table B1.

Manipulating Age — The age of the startup founder is indicated by the graduation year from their college or graduate school rather than being listed directly.[36] If a team has two co-founders, their age falls in the same range, which belongs to either the older group (who graduated before 2005) or the younger group (who graduated after 2005). I assume founders graduate from college at the age of 23,[37] so the approximated age is calculated by the formula: age $= 2020 - \text{graduation year} + 23$.

Evaluation Questions

The evaluation questions include three mechanism questions, designed to directly test belief-based sub-mechanisms, and two decision questions, which were designed to compare investors' initial contact interest and later stage investment interest. Considering that most venture capitalists are well-educated and market savvy, I use probability or percentile

[33] Crunchbase is a commercial platform that provides public information about startups mainly in the U.S.

[34] I do not provide private information like equity sharing plans because such information is generally not disclosed to investors in the pre-selection stage.

[35] Considering our collaborative incubators and startups have relatively more Asian founders and female founders, the ratio of female and male startup founders are both 50% to maximize the experimental power. A similar ratio is used for Asian founders and white founders.

[36] It is suspicious to list age directly in a startup profile because it is not common practice. Hence, using the graduation year as a proxy achieves more realism.

[37] Using 22 gives similar results. However, some investors may assume the founders graduate from graduate school rather than from an undergraduate program at these universities; hence, 23 is used.

ranking questions rather than Likert scale questions.[38] This design has two advantages. First, probability or percentile ranking questions are more objective. Second, the wide range from 1 to 100 provides richer evaluation results and additional statistical power. This allows researchers to implement infra-marginal analysis and distributional analysis that explore how investor preferences change across the distribution of contact and investment interest. Screenshots showing the appearance of these questions are provided in Online Appendix B Figure B7, and Figure B8.

(Belief-based) Mechanism Questions — The three mechanism questions are designed to test the following three standard, belief-based discrimination mechanisms. First, being a minority can be indicative of the startup's future financial returns. To test this mechanism, investors need to evaluate the percentile rank of each startup profile compared to the startups they have invested in previously, which is the quality evaluation question ($Q1$). Second, investors may worry that minority founders are willing to collaborate with other types of investors or to use other financial tools for their fundraising purposes.[39] To test this channel, investors need to evaluate the probability that the startup will accept their investment rather than other investors', which is the "availability" evaluation question ($Q2$). Third, investors may use the founder's group membership as indicative of the startup's risk (i.e., the second moment). Therefore, investors also evaluate the risk percentile rank of each startup profile compared with the startups they have invested in previously, which is the risk evaluation question (Q_5).[40] This risk evaluation question is added when I recruit investors using only the matching incentive for robustness test purposes.[41]

Decision Questions — The correspondence test generally only observes the initial contact interest of candidate evaluators rather than any later stage decisions. To understand more of VC's investment criteria, the two decision questions are designed to examine both the investor's initial contact interest ($Q3$) and later stage investment interest ($Q4$). The investment interest question asks the relative investment interest rather than the investment magnitude mainly because different investors have different ranges of targeted investment amounts. In order to accommodate more investors, I try to make the question as standardized and generally applicable as possible.

Background Questions

[38] Similarly, Brock and De Haas (2020) use probability questions to replace Likert Scale questions when they recruit real Turkish bankers to evaluate different loan profiles.

[39] Similar to the marriage market, the entrepreneurial financing process is also a two-sided matching process. Therefore, this type of "availability" potentially also matters.

[40] For special characteristics like founder's gender, race, and age, the first mechanism question (Q_1) tests one of the most common statistical discrimination mechanisms. The second mechanism question (Q_2) tests a typical confounding mechanism in a two-sided matching market in the discrimination literature. The third mechanism question (Q_5) sheds light on whether the belief of expected variance affects an investor's decision, which is discussed in detail in Neumark (2012) and Heckman (1998).

[41] During the recruitment process, I received feedback to add this question from several investors. Therefore, when recruiting the rest of the investors using only the matching incentive, a risk evaluation question was added at the end of all the evaluation questions to minimize its impact on all the other questions while collecting information about this mechanism. An alternative way to obtain such information is to implement a new field project (for example, send an extra survey) as done by Bartoš et al. (2016).

After the investor evaluates all the profiles, I also collect standard background information about the participant to check how representative my sample investors are and analyze potential heterogeneous effects based on predetermined investor characteristics.[42] Such background information includes investors' preferred industries, stages, special investment philosophies, gender, race, and other standard demographic information.

B.2 Donation Section (Dictator Game)

To directly identify the taste-based mechanism, I insert a donation section at the end of the survey tool. Investors will be informed that they will receive an *unexpected* $15 Amazon Gift Card to thank them for participating in this experiment.[43] However, they can decide whether to donate a portion of the provided $15 to the displayed startup founders. For instance, if the investor donates $3, she/he will receive a $12 Amazon Gift Card. The researcher will use the donated money to purchase a small gift for the corresponding type of startup founders in our collaborative incubators and bring them anonymous encouragement.[44] The detailed donation question and the example founder's picture are provided in Online Appendix B Figure B13.

I randomize the gender and race of the startup founders receiving the small gift by changing the pictures displayed and the wordings used in the description. The options investors may randomly be provided with include donating to the "Women's Startup Club" (mainly white female founders), "Asian Women's Startup Club" (mainly Asian female founders), "Asian Startup Club" (mainly Asian male founders), or just "our Startup Club" (mainly white male founders). To make the information of gender and race more salient, I also add a picture containing four startup founders of the same gender and race so that survey participants understand what type of founders they are donating to.[45] All individuals in the pictures are smiling and professionally dressed to make sure they are as much on equal footing as possible.

Incentive Structure As a preference elicitation technique, one key point of the IRR experimental design is its incentive structure. The following incentives are designed not only to increase the stakes of Experiment A and impose costs for making inefficient and inaccurate evaluations, but also to bring real value to all the experiment's participants. The related incentive structure for each evaluation question is provided in Online Appendix B Table B4.

[42]It is important to ask the background questions after the evaluation section in order to avoid priming subjects to think about any particular characteristics that the research project aims to test.

[43]To avoid polluting the incentive structure in Experiment A, the compensation with the $15 Amazon Gift Card is mentioned only at the very end of the survey tool rather than in the consent form.

[44]The reason why I provide a small gift rather than cash to founders is that a small gift is usually more associated with warm encouragement. Giving a small amount of cash can be insulting to someone.

[45]The concern of using pictures in the experiment is that the appearance or other messages delivered by the pictures cannot be fully controlled. To ameliorate this issue, I use four founders' pictures combined together to send the signal of gender and race. All the pictures are obtained from a public library (i.e., Wikimedia Commons, Freeimages, etc.) with no copyright problems.

Matching Incentive — For the randomly selected 4,000 investors who receive the recruitment email (Version 1), I only provide the following "matching incentive" following Kessler et al. (2019). Basically, after each investor evaluates all the startup profiles, we use a machine learning algorithm to identify matching startups from our collaborative incubators. If the matched startups are also interested in the investor's investment philosophy, they will contact the investor for a potential collaboration opportunity. The matching algorithm uses all of their evaluation answers to identify their investment preferences. Therefore, all five evaluation questions are incentivized by this incentive. I provide a description of the algorithm in the consent form, and the details of the algorithm are provided in Online Appendix B.[46]

Monetary Incentive — To increase the sample size, I provide both the "matching incentive" and an extra "monetary incentive" used by Armona et al. (2019) to the other randomly selected 14,000 investors who received the recruitment email (Version 2). This "monetary incentive" is essentially a lottery in which 2 experiment participants will be randomly selected to receive \$500 each plus an extra monetary return closely related to their evaluations of each startup's quality. The more accurate their evaluations are of each startup's quality, the bigger the financial return they will obtain as a lottery winner.[47] However, this is only used to incentivize the "quality evaluation question" (i.e., "Q1"). The evaluation results will be determined based on the Pitchbook data published in the next 12 months after the recruitment process is finished. I separately informed these two lottery winners that they would receive the award at the end of July 2020. The evaluation algorithm is provided in the consent form (Version 2).[48]

Justification — One concern of adding the "monetary incentive" is the possibility of attracting participants who do not value the matching incentive, which results in extra noise. I have compared the evaluation results of investors who receive only the "matching incentive" and those who receive both incentives. The comparison results are provided in Online Appendix B Table B6, showing that this potential extra noise does not cause systematic different evaluation results. The interaction terms between the incentive structure and startup's gender and race are not significant. Moreover, this experiment discovers multiple highly significant startup team and project characteristics that are crucial

[46]This "matching incentive" has the following merits. First, researchers can apply it to any other two-sided matching markets. Second, it can incentivize all the evaluation questions compared with the monetary incentive. Third, if the designed matching algorithm can improve the matching efficiency, such an incentive can bring real value to both sides of the matching market. Despite the merits mentioned above, implementing this incentive often requires researchers to have certain social resources and connections.

[47]For example, Peter Smith, one experiment participant, is chosen as the lucky draw winner. In his survey, he indicates that on average, he feels that male teams are more likely to generate higher financial returns. In that case, we then construct a portfolio containing more real startups with male teams. After one year, based on the financial performance of real startups, this portfolio containing more startups with male teams generates a 10% return. So Peter Smith receives \$500 + \$500*10% = \$550 as his finalized monetary compensation one year after he participates in the survey. \$500*10%=\$50 is the "extra monetary return." The historical return of the VC industry is roughly between -15% and +15%, which means that the range of expected monetary compensation is roughly between \$425 and \$575.

[48]This "monetary incentive" has both merits and limitations. First, it mimics the real investing process in which investors have a certain amount of principal and need to evaluate different startups accurately to generate maximum return. Second, it does not require many social resources. Third, researchers can apply it to more general situations besides a two-sided matching market. However, the current version can only incentivize the evaluation of startup's quality (i.e., Q1) to avoid distorting participants' evaluations on other questions. If the collaboration likelihood (i.e., Q2) is added to the financial return algorithm, then all the participants may claim that the best startups would be willing to collaborate with them even if it is not true. Similarly, if contact interest (i.e., Q3) and investment interest (i.e., Q4) are added to the financial return algorithm, participants may be motivated to distort their true evaluation in order to maximize their financial return as both Q3 and Q4 can be affected by Q2.

for investors' investment interest. This means that investors understand the incentives and evaluate all the questions carefully.[49]

3.2 Experimental Results

A. (IRR Experiment) No Group-level *Explicit* Bias Against Minority Founders.

I denote an investor i's evaluation of a startup profile j on evaluation question k as $Y_{ij}^{(k)}$ and estimate the following regression, which allows me to investigate the average response to a founder's demographic information across recruited investors in our study. Formally,

$$Y_{ij}^{(k)} = X_{ij}\beta^{(k)} + \alpha_i + \epsilon_{ij}^{k} \tag{1}$$

X_{ij} represents any founder's demographic information, like gender, race, or age. α_i are investor fixed effects that account for different average ratings across investors. Since each type of startup characteristic is randomized orthogonally and independently, the coefficient $\beta^{(k)}$ has a causal interpretation.

Table 4 reports regression results testing group-level *explicit* bias based on founder's gender (Panel A), race (Panel B), and age (Panel C) by using the total 1,216 profile evaluation results. Female (Asian) Founder is a dummy variable that is equal to one if the startup founder has a female first name (Asian last name), and zero otherwise. Age is the approximated founder's age based on the graduation year from the college. The dependent variables are investors' evaluation results for startup's quality ($Q1$), willingness to collaborate ($Q2$), risk ($Q5$), and investors' interests in contacting the team ($Q3$) and investing in the team ($Q4$). All the regressions include the investor fixed effect to control for any subjective judgement by each individual investor.

The regression results show that most of the coefficients on founder's gender, race, and age are not significantly different from zero, indicating that there is no group-level *explicit* bias against minority founders.[50] This null result potentially stems from the following reasons. First, investors understand that they are participating in a research project and hence behave more friendly to minorities. Second, the recruited investors are more pro-social, leading to sample selection bias. Lastly, founder's group membership is not the first-order characteristic that profit-driven investors care about. The first two reasons imply that the detected *explicit* bias serves as the lower bound of the true

[49] See Online Appendix B Table B5. Researchers can also separately ask subjects questions that can test their understanding of the incentive (Casaburi and Willis (2018)).

[50] The only exception is that older founders are considered to be slightly less risky at the 0.10 level of significance. However, given that only a small number of investors need to answer $Q5$, I do not focus on this result. Also, I do not find investors spend significantly different amounts of time on evaluating majority founders' and minority founders' profiles.

bias in the real world.

B. (IRR Experiment) Group-level *Implicit* Bias Against Minority Founders. (Belief-driven)

According to Bertrand et al. (2005b) and Cunningham and de Quidt (2015), implicit bias can significantly affect people's behaviors when the task involves ambiguity, evaluating mixed attributes, or people face time pressure. The task of screening startups in the pre-selection stage satisfies these criteria. Specifically, many investors need to evaluate a large number of startups quickly before narrowing down their potential investment targets. Hence, it is important to test whether implicit bias exists. Table 5 reports regression results testing group level *implicit* bias based on founder's gender (Panel A), race (Panel B), and age (Panel C). Following Kessler et al. (2019), I check whether investors' ratings of minority founders decline when they feel fatigued, which requires me to compare investors' evaluations before and after the inserted short break (i.e., compare evaluations in the first half and second half of the study). "Second Half of Study" is an indicator variable for startup profiles shown among the last eight profiles viewed by the investor. In column (1), the dependent variable is investors' response time, defined as the number of seconds before each page submission and winsorized at the 95^{th} percentile (59.23 seconds on average).

Regression results prove the existence of the group-level *implicit* bias against female and Asian founders.[51] Column (1) shows that investors indeed become fatigued or rushed in the second half of the study as they spent 27 seconds fewer evaluating each profile after the inserted break. Columns (2) and (5) indicate that investors' "quality ratings" and "investment interest" of female (Asian) founders' startups gradually decline compared with similar male (white) founders' startups in the second half of the study, respectively. In Columns (2)-(5), all the interaction terms of "Female Founder" ("Asian Founder") and "Second Half of Study" are negative. These interaction terms are significantly negative for Columns (2) and (5), suggesting belief-driven *implicit* bias against female and Asian founders.[52] However, older founders do not suffer from this implicit bias. Panel Cs of Table 5 and Table 4 even show that investors have both "explicit" and "implicit" belief that older founders' startups are less risky.

Alternative Interpretation — Learning Effect One alternative interpretation of the "fatigue effect" is that investors' ratings of minorities will decline when they are more familiar with evaluating these profiles. This indicates a worse situation where investors' biases are essentially "explicit", especially when they are more experienced. After I remove the first few evaluations of each investor, the "fatigue effect" still holds, indicating that "learning" may not

[51] Results are not significant for older founders. Also, the evaluation time does not significantly decrease for minorities in the second half of the study, ruling out attention discrimination in the profile screening process (see Appendix B Table B9).

[52] Although the p-values of "Asian Founder" in the second half of study are about 0.2, Table 7 later shows that in "high contact interest" situations, there is a strong implicit bias against Asians.

be the main reason. However, if investors are still learning until the end of the evaluation section, then I cannot fully rule out this channel due to the lack of power.

Rule Out Other Alternative Interpretation Another alternative interpretation for the discovered "fatigue effect" is that investors may "expect" the overall population of entrepreneurs to follow the distribution in the real world, say "80-20", rather than "50-50" as used in this experiment.[53] Then near the end of the reviewing process, investors may be tempted to skew the sample in the direction of what they expect the population to be, leading to biased experimental results.[54] To test this hypothesis, I check whether investors' evaluation results are influenced by the "mixed profiles" they have already evaluated in the first half. However, I do not find systematic evidence that the more minority founders the investor evaluated in the first half, the tougher she/he is on the minorities in the second half. Results are provided in Online Appendix B Table B7.

B.1 Heterogeneous Effect Based on Investors' Sectors

Table 6 tests implicit gender bias for investors working in tech sectors and non-tech sectors. According to the literature discussing gender issues in sciences or STEM industries (Carrell, Page and West (2010), Goldin (2014), and Kessler et al. (2019)), investors from tech sectors potentially have more implicit bias against female founders. Table 6 Panel A focuses on tech-sector investors mainly working in the Information Technology (IT) industry. Columns (2) and (5) show that these tech investors have stronger belief-driven implicit bias against female founders based on their quality evaluations and indicated investment interest. However, results of Panel C show that non-tech investors working in other industries, such as the education industry, media industry, or entertainment industry, do not have this implicit bias against women.

To help understand the magnitude of this implicit gender bias in the tech sector, Table 6 Panel B focuses on the second half of the study, and uses the effect of going to an Ivy League College as the benchmark to calculate the relative magnitude of this implicit bias. Panel B Column (1) shows that on average, going to an Ivy League College increases investors' quality evaluations by 8.78 percentile rank. Then the implicit gender bias accounts for 44% of the Ivy League College effect. Columns (4) and (5) show that for investors' contact interest and investment interest, this implicit gender bias accounts for roughly 37% and 60% of the Ivy League College effect respectively. The magnitude is not trivial, suggesting that the implicit gender bias could play an important role in tech sector investors' decisions.

[53] "50-50" is chosen to maximize the experimental power. However, using this ratio may also keep reminding investors of the experiment's purpose (i.e., testing discrimination), making investors behave more friendly, especially in the second half where the balanced ratio is more obvious.

[54] Special thanks go to Peter DeMarzo who raised this brilliant point and suggested the corresponding testing method.

B.2. Distributional Effect (Reconcile the Literature)

As discovered in Zhang (2021) and also pointed out by Lundberg and Startz (1983), Neumark (2012), and Heckman (1998), the magnitude and direction of evaluators' socially sensitive preferences can vary across the distribution of candidates' ability or quality. Therefore, to reconcile mixed results in the literature, this section examines how investors' implicit gender, racial, and age bias varies across startups' quality. Since correspondence tests and Ewens and Townsend (2020) only observe the contact stage, I need to use investor's likelihood to contact the startup (i.e., "Q3 contact interest") rather than the investment interest (i.e., "Q4") to reconcile the mixed results in the literature.[55]

Figure 6 demonstrates these distributional effects by analyzing the profiles evaluated in the second half of the IRR experiment.[56] Panels A, C, and E provide the empirical cumulative density function (CDF) for a founder's gender, race and age across investors' contact interest ratings, respectively. Panels B, D, and F provide the OLS coefficient estimates and the corresponding 95% confidence intervals for a founder's gender, race, and age across investors' contact interest ratings, respectively.

The distributional effects discovered in Figure 6 show that both the direction and magnitude of investors' implicit gender, racial, and age bias vary across the startup's quality. When evaluating "low-quality" startups which receive less contact interest, investors are biased *towards* female, Asian, and older founders.[57] However, when evaluating "high-quality" startups which receive more contact interest, investors are biased *against* these minority founders. For example, based on Panel A, when investors' contact interest ratings are lower than 5%, the CDF for a female founder is slightly to the right compared with the CDF for a male founder. In Panel B, the coefficient of being a female founder is also slightly positive on the left tail of the distribution. All of these indicate that among "low-quality startups", investors slightly prefer female founders. However, as investors' contact interest increases, this bias reverses and investors start to prefer male founders. Similar "discrimination reversion" phenomenon also exists for racial and age bias as shown in Panels C,D,E, and F.

Motivated by Figure 6, Table 7 confirms this "discrimination reversion" phenomenon using regression methods. Panel A shows that investors' implicit racial bias mainly exists when evaluating "high-quality" startups (i.e., $Q3 \geq 50\%$).[58]

[55] There are two key differences between investment interest ("Q4") and contact interest ("Q3"). First, investment interest ("Q4") is noisier than contact interest ("Q3") because this experiment does not provide any soft information about the startup founder, which is also crucial to investment decisions. Second, the investment interest is mainly affected by beliefs rather than tastes as documented by Zhang (2020). Therefore, when using investment interest to illustrate the distributional effect (see Appendix B Figure B4), the discrimination reversion pattern is slightly noisier for Asian founders and older founders, and it disappears for female founders. The reason is because any taste-driven preference towards women (proved in Experiment B) mainly plays a role in the contact stage rather than the investment stage. These results are expected according to Zhang (2020).

[56] Figure B3 in Appendix B uses the total profile evaluations. Although the magnitude of bias is much smaller due to the potential consent form effect, patterns are still similar to that in Figure 6.

[57] Older founders are defined as founders who graduated from college before 2005.

[58] Results are still robust if $Q3 \geq 40\%$ or $Q3 \geq 60\%$.

Columns (2), (4), and (5) show that in the second half of the study, Asian founders receive significantly lower ratings than white founders in profitability evaluations (Q1), contact interest (Q3), and investment interest (Q4). Panel B discusses the magnitude of this bias. Columns (2), (4), and (5) show that the implicit bias against Asian founders accounts for 49%, 38% and 80% of the Ivy League College effect for investors' quality evaluation, contact interest, and investment interest, respectively. However, results of Panel C show that when evaluating "low-quality" startups (i.e., $Q3 \leq 50\%$), investors do not have this implicit bias.

The discovered "discrimination reversion" phenomenon provides a crucial insight to reconcile the contradictory results of previous literature. On the one hand, papers exploiting regression-based methods often use observational data of relatively high-quality startups, which either have successfully raised funding or are mature enough to post information on large fundraising platforms (Ewens and Townsend (2020)). Hence, these papers mainly capture the middle and right part of the startup quality distribution where investors are more biased *against* female and Asian founders. On the other hand, papers using the correspondence test method often focus on the cold call pitch email setting. The low response rate to these cold call emails shows that this experimental method mainly captures the left tail of the startup quality distribution where investors have less contact interest. Not surprisingly, correspondence tests find that investors are slightly biased *towards* female and Asian founders. To sum up, previous literature generates mixed results because they capture investors' preferences for startups in different parts of the startup quality distribution. When discrimination can reverse across this distribution, results from the correspondence test cannot be generalized to more mainstream fundraising settings.

To fully understand why investors' bias can reverse, researchers need to identify the nature of such bias across the startup quality distribution. Experiment A shows that the implicit bias against female and Asian founders of high-quality startups is mainly driven by the belief that minorities' startups are less profitable. In section 4, Experiment B will identify the nature of investors' bias when they evaluate low-quality startups.

B.3. Decision-based Heterogeneous Effect and "Leave-one-out" Estimator

In the real world, investors' preferences towards the founder's gender, race, and age are heterogeneous. Some are biased towards minority founders, while some are biased against minority founders. In a divided society, different groups potentially make opposing decisions based on different motivations. For example, "pro-minority" investors' investment decisions can be driven by taste if they simply want to support minorities rather than maximize financial returns. However, "anti-minority" investors' investment decisions can be driven by belief if they believe minority founders' startups are less profitable. Understanding the separate driving forces of these two different groups' decisions has important policy implications. To test what separates investors and how divided the investment community is, I develop a consistent, "*decision-based*" heterogeneous effect estimator using the "leave-one-out" technique.[59]

The logic behind how this estimator works is very simple. The IRR experiment introduces "*within-individual*" level randomization and requires investors to evaluate multiple randomized startup profiles. Therefore, when the number of profiles evaluated by each investor is large enough, researchers can identify "individual-level" preferences based on investors' contact/investment decisions in an ideal setting. Theoretically speaking, it is feasible to accurately tell which investor is biased against minorities and classify recruited investors into a "pro-minority" group or an "anti-minority" group. Researchers can then run separate pooled regressions within each group to investigate each group's mindset. However, in a nonideal situation, creating a consistent estimator requires solving the potential generated regressor problems by using the "leave-one-out" technique. Detailed proof and discussions of this estimator are provided in Appendix E.[60]

Table 8 provides the decision-based heterogeneous effect for founders' gender, race, and age, which measures the evaluation results of "pro-minority" investors and "anti-minority" investors. Panels A, B, and C report the contact decision-based heterogeneous effect of startup founder's gender, race, and age by using the second half of evaluation questions. All the coefficients and standard errors in the parentheses are calculated using the "leave-one-out" estimator and the bootstrap method due to the relatively small sample size.

Results of Table 8 Panel A show that "pro-women" investors and "anti-women" investors have very different expectations of women-led startups' profitability. Column (1) shows that "anti-women" investors feel that women-led startups have 16.40 percentile ranks lower potential financial returns than men-led startups. They also have lower contact interest and investment interest for female founders compared with similar male founders. However, "pro-women" investors expect that women-led startups have 7.93 percentile ranks higher potential financial returns than men-led startups, and they also prefer contacting female founders than male founders. Panel B and Panel C show that for "anti-Asian" and "anti-older" groups, lower expectations of these founders' profitability is also the important reason why investors do not want to contact Asian and older founders. Similar to the gender bias, the split in profitability expectations for Asian-led and older-led startups is one explanation for investors' divided decisions.[61] Moreover, the

[59]Traditionally tested heterogeneous effect often relies on subjects' predetermined demographic information. For example, researchers separately test the investment preferences of well-educated investors and uneducated investors. However, the "*decision-based*" heterogeneous effect estimator classifies the investor pool into a "pro-minority" group and an "anti-minority" group based on each investor's indicated decisions.

[60]It should be noted that the current version of this estimator still relies on the assumption of linearity and the more generalized form of this estimator will be provided in future papers.

[61]However, as shown in Section 3.2, "pro-older" investors' decisions are also affected by the belief that older founders impose less risk. For "pro-Asian" groups, the donation results will later show that taste also plays a role as our sample investors are more friendly to Asian founders.

division in investors' attitudes towards female founders is larger than that towards Asian and older founders.

Do women-led startups perform worse than men-led startups during the COVID-19 period? To examine the accuracy of investors' beliefs, I compare the performance of women-led startups and men-led startups between 07/2020-07/2021/ (i.e., during the one-year period after the experiment).[62] Online Appendix B Table B8 shows that conditional on prior funding, women-led startups generally perform similarly to men-led startups in terms of the likelihood of raising new funding, going out of business, or successfully exiting through IPO or M&A. However, women-led startups underperform in the IT industry and specifically are significantly associated with a higher likelihood of going out of business in this industry. This, to some extent, justifies investors' gender bias against women as investors need to screen out more women-led startups in the pre-selection stage to achieve a similar financial performance to investing in men-led startups.[63] It should be noted that the under-performance of women-led startups may be a temporary phenomenon due to COVID-19.

C. (Dictator Game) Detect Taste-Based Gender Homophily

Table 9 reports the regression results from the dictator game in which investors' *anonymous* donation decisions do not affect their investment opportunities or help improve investors' social images. Hence, results only test the potential taste-driven bias based on the startup founder's gender and race. The dependent variable is the donated amount measured in dollars, ranging from $0 to $15. In Columns (1)-(3), I include the investors who did not select a donation amount and treat their behaviors as "donate 0$". In Columns (4)-(6), I exclude the investors who did not select the donation amount. "Female Founder" and "Asian Founder" are indicative variables that are equal to one if the displayed startup founder is female and Asian, respectively, and zero otherwise. Similarly, "Female Investor" and "Asian Investor" are indicative variables that are equal to one if the investor is female and Asian respectively. All regressions use robust standard errors reported in parentheses.

Columns (1) and (4) show that investors are biased towards Asian founders, which makes sense because these experiment participants are likely to be more pro-social. Columns (2) and (5) show that male investors, on average, donate $3 less to female founders compared to similar male founders, indicating a weakly significant group-level, taste-driven bias against female founders. However, Column (5) shows that female investors are significantly more likely to donate money to female founders even when the support cannot be observed by founders and they will sacrifice the real mon-

[62] Bordalo, Coffman, Gennaioli and Shleifer (2016) and Bordalo, Coffman, Gennaioli and Shleifer (2019) have detailed discussions about the implications of incorrect beliefs. The method used here is similar to the "outcome test" mentioned by Becker (1993).

[63] Barber, Jiang, Morse, Puri, Tookes and Werner (2021) show that research productivity falls more for women during the COVID-19 pandemic. I do not check the performance of Asian and older founders' startups because information about founders' race and age is not well-recorded on Pitchbook.

etary award. The results are consistent with the homophily channel that female investors are more likely to support female founders. However, consistent with DellaVigna et al. (2013), I find that men are usually more generous than women during the donation process. In terms of the absolute amount of money donated to female founders, male investors donate \$9.36 while female investors donate \$9. It is just that male investors are even more generous to male founders, donating \$12.41 to them.[64]

3.3 Discussion

Experiment A has the following advantages. First, it provides a strong incentive to reveal investors' preferences by providing real investment opportunities. This incentive structure also brings real benefits to all the experiment's participants. Second, by combining the IRR experiment and the dictator game, Experiment A can directly test multiple belief-driven mechanisms and taste-driven mechanisms. Third, carefully designed evaluation questions can shed light on investors' later stage decisions and allow the testing of decision-based heterogeneous effects. Lastly, by obtaining rich evaluation outcomes within each individual, this experimental design generates a large enough sample size from a limited number of participants. This enables its feasibility in a wide application of financial studies whose experimental subjects are usually busy investors.

Despite all the merits, it is helpful to note its limitations. First, similar to any experiments requiring voluntary participation, Experiment A also has potential sample selection bias during the recruitment process. This does not hurt the internal validity of the experiment, but it is helpful to check the external validity by running a complementary experiment or replicating this experiment in different countries. Second, the consent form potentially could affect participants' behaviors due to the observer effect. This implies that any detected bias is likely to be the lower bound of the existing bias in the real world. Using appropriate incentive structure and de-identified data can ameliorate this issue. Third, Experiment A's incentive structure requires many social resources, so any innovation on providing cheaper incentive structures is imperative to lower the experiment's cost. Lastly, Experiment A does not observe investors' real investment amounts. Hence, we also need studies exploiting "natural experiments" in the future.

To test the external validity of Experiment A and fully understand why investors can be biased towards minorities, I follow up with Experiment B. Experiment B uses a redesigned correspondence test to identify the nature of discrimination in the cold call, pitch email setting. Also, before the academic community widely uses the IRR experimental design in the future due to its various powerful functionalities in testing discrimination, it is interesting to compare its results with the currently widely used correspondence test.

[64] Homophily is also documented by Raina (2019).

4 Experiment B's Design and Results

Experiment B studies gender and racial discrimination with 17,000+ real venture capitalists mainly from the U.S. and other English-speaking areas in the world.[65] To test the nature of any detected biases, Experiment B utilizes a new email technology that tracks investors' detailed information acquisition behaviors. This enables Experiment B to introduce meaningful variation in startup quality in addition to randomizing the startup founder's gender and race.[66]

4.1 Experimental Design

A. Investors and Recruitment

The correspondence test was mainly implemented between 03/2020 - 04/2020 during the outbreak of COVID-19. During this period, President Trump started using the expression "Chinese Virus" on 03/18/2020 and later stopped using it around 03/23/2020 to protect the Asian American community.[67] Accidentally, this correspondence test captures this special "Chinese Virus" period.[68] Investors recruited for Experiment B are mainly early stage venture capitalists in the U.S. and other English-speaking areas in the world as documented in Section 2. Table C4 in Online Appendix C provides the industry distribution of the created hypothetical startups. In total, I create 67 startup ideas and use more than 200 names to make sure that all experimental results are not driven by any special names or startup ideas.[69]

B. Randomization and Design

Manipulating Identity of the Entrepreneur. — I assign four co-founders for each created startup team, which include a white female co-founder, a white male co-founder, an Asian female co-founder, and an Asian male co-founder.[70] Each co-founder has a randomly assigned first name and last name that signal their gender and race. To make sure that investors associate the names with the correct gender and race information, I have recruited 107 U.S.-based Amazon Mechanical Turk users to assess the gender and race of these selected names to delete any ambiguous names. The

[65] These areas include the UK, Canada, Australia, Singapore, Hong Kong, Isreal, India, etc.

[66] To test the nature of discrimination, researchers need to randomize startup characteristics that affect investors' contact interest. In Experiment B, I orthogonally randomize the startup founder's gender, race, educational background, and the project's comparative advantages.

[67] See https://theconversation.com/donald-trumps-chinese-virus-the-politics-of-naming-136796 and see Forbes News "Trump Abruptly Stops Calling Coronavirus 'Chinese Virus' At Daily Press Briefing".

[68] Considering the unusualness of this period, I implemented another round of the correspondence test on the same pool of investors in 10/2020. However, the second-round experiment provides very noisy results as many investors have realized the existence of this experiment. Hence, this paper mainly shows results from the first round. Results of the second-round experiment are available upon request.

[69] These ideas cover the majority of mainstream industries that venture capitalists are interested in, which include Information Technology, Healthcare, Consumers, Energy, etc.

[70] Having co-founders for a startup is very common, especially for highly innovative and complicated companies. Based on Pitchbook data, startups with multiple co-founders account for 50% of all startups.

name lists and the name generation process details are provided in Appendix C.

Manipulating the Startup's Quality — I randomize the startup team's educational background and the project's comparative advantages in both the subject line and the contents of each email. For the educational background, the control group does not mention the founders' educational backgrounds. However, the treatment group indicates that the startup team comes from a prestigious university in the U.S. in both the email's subject line and contents.[71] Similarly, for the project characteristics, the control group does not mention any specific comparative advantages of the startup while the treatment group mentions comparative advantages such as "22% MOM Growth Rate" or "Patent Registered."[72]

Pitch Email Design and Website Construction—The pitch emails, covering the 67 startup ideas written for this experiment, follow the template and structure provided by Gornall and Strebulaev (2020a) and "good pitch email template" examples posted on Quora. The startup ideas are provided by my research team members, who are usually young startup founders or members of startup-related clubs at Columbia and other Ivy League colleges.[73] We use Wix, a commercial website builder, to make the related startup websites which are in the under-construction stage. The pitch email example is provided in Figure 3, and the website example is provided in Figure 4.

Manipulating Access to Information. —The randomization of startups' characteristics is implemented in the following two stages. For the first stage, before the investor opens the pitch email, she will see the randomly assigned email sender's name indicating the sender's gender and race,[74] and also the randomly generated email subject line indicating whether the startup has a well-educated founding team and a project with an impressive advantage. For the second stage, after the investor opens the pitch email, she will decide how much attention to spend reading this pitch email. In each email's contents, the co-founder's name occurs multiple times (including in the introductory paragraph, email addresses, the email signature, and email senders' names) to make the gender and race information more salient. If the email's subject line mentions an Ivy League educational background or project advantages, there are extra sentences inserted to emphasize this information again in the email's contents while keeping the rest of the contents the same. After reading the email's contents, the investor can decide whether to reply or forward the email to other related investors who may potentially also be interested in the same pitch email. All the technical details about

[71] Prestigious universities used include Ivy League colleges, MIT, and Stanford. In the experiment implemented between 03/2020 and 04/2020, I also included Northwestern University, Caltech, Johns Hopkins University, Juilliard School, and other top schools in the field related to the startup. For example, if the startup is related to music, I mention that the founding team members come from Columbia University and the Juilliard School.

[72] MOM is an abbreviated form for "month over month" growth.

[73] We only choose the valid startup ideas with relatively good coverage of key industries after discussions with practitioners.

[74] Although large companies may ask the secretary or investor relationship manager to contact investors, for early-stage startups, it is usually the startup's founding team members themselves who contact investors in order to show their sincerity.

sending a large number of emails and the preparation work are provided in Online Appendix C.

Email Behavior Measurements — I track the following email behavior measurements: the email opening status and the corresponding time stamp, the email staying time measured in seconds, the sentiment of email replies analyzed through LIWC, the click rate of the inserted startup websites, the response rate and contents of replies.[75] Despite these rich behavior measurements, only email opening rate and email staying time generate enough power to analyze investors' responses. All the other traditionally used behavior measurements do not survive in the recession period when the "low-response-rate" problem is more severe than before. The detailed mechanisms of recording different email behaviors and whether such behavior measurements are used in previous literature are described in Online Appendix C Table C5. The flow chart of the first correspondence test is provided in Figure 5.

4.2 Results

4.2.1 *Bias towards Female and Asian founders in the Pitch Email Setting.*

Table 10 Panel A summarizes investors' major information acquisition behaviors in the correspondence test. On average, the pitch email opening rate is 12.03% and each investor spends roughly 24 seconds reading the cold call pitch email in 03/2020-04/2020. However, both the startup website click rates and the email response rates are very low (roughly 1%), indicating that early-stage investors are sensitive to business cycles as documented by Howell et al. (2020). Therefore, traditionally recorded investors' email behaviors, such as the email response rate, do not generate enough experimental power during the COVID-19 pandemic. All the experimental results in this paper rely on the new email behaviors recorded by the latest email tracking technology.

Table 10 Panel B reports regression results of global investors' email opening behaviors for randomized pitch emails. The dependent variable is a dummy variable, which is one when an investor opens the pitch email, and zero otherwise. Columns (1), (2), (3), and (5) use all the observations collected during 03/2021-04/2021. In Column (4), results are reported for the sub-sample where the startup team's educational background is from "purely Ivy League colleges", Stanford, and MIT.[76] All the regressions include start-up fixed effects to control for any idiosyncratic characteristics of each start-up pitch email, such as the business models, etc. Hence, I am comparing investors' email opening rates within the same start-up's pitch email, and all the results are similar after including investor fixed effects. Following Bernstein, Korteweg and Laws (2017), all standard errors are clustered at the investor level to account for the corre-

[75] LIWC (Linguistic Inquiry and Word Count) is a text analysis program used for sentiment analysis.

[76] "Pure_Ivy" represents cases like "Team from Columbia University," while "Mixed_Ivy" represents cases like "Team from Columbia University and Juilliard Music School." For some startups in the music or medical industry, I combine an Ivy League college with a good university in that specific area for the treatment group.

lated opening decisions across different pitch emails received by the same investor.

Results of Table 10 Panel B show that on average, investors are biased towards female, Asian, and well-educated founders during the experimental period. Column (1) shows that using a female first name in a pitch email raises the opening rate by 1% compared to using a male first name. This difference is statistically different from zero at 1% significance level. Considering that the base opening rate is 12.03% for all the pitch emails, this represents an 8% increase in opening rates. Column (2) shows that using an Asian last name in a pitch email raises the opening rate by 0.7% after President Trump stops using the wording "Chinese Virus" in 03/2021.[77] This difference is statistically significant at 10% and represents a 6% increase in opening rates compared with using a white last name. Similarly, Column (3) shows that mentioning a good educational background in the email's subject line increases the opening rate by 0.7%. This effect increases to 1.2% if I focus on the sub-sample with only a pure Ivy League educational background mentioned (i.e., "Team from Columbia University" rather than "Team from Columbia University and Juilliard Music School "). This represents a 10% increase in email opening rates compared with not mentioning anything in the email's subject. However, Columns (4) and (5) show that mentioning the project advantages in the email's subject line does not significantly increase the email opening rate. Results provided by Table 10 confirm the surprising results found in previous literature that investors are biased towards female and Asian founders in the pitch email setting.

4.2.2 *The Direction of Bias Depends on Timing.*

Table 10 Panel C reports regression results of how startup characteristics affect investors' staying time on each email. The dependent variable is the time spent on each pitch email measured in seconds, which approximates how much attention each investor spent. In Columns (1) and (2), I include unopened emails and replace their email staying time with 0 seconds. Considering the potential truncation issue, I also report the sub-sample of opened emails in Column (3).[78] Similar to Panel B, all the regressions include start-up fixed effects, and standard errors in parentheses are clustered at the investor level.

Results of Table 10 Panel C show that although investors generally spent more time on female and Asian founders' emails, there was a temporarily strong bias against Asian founders during the COVID-19 outbreak in 03/2020. Column (1) shows that using a female first name raises the time spent on a pitch email by 0.36s in 03/2020 and 0.12s in 04/2020. This magnitude is not large due to the truncation issue. Similarly, Column (3) shows that using an Asian

[77] See Trump Says He'll Stop Using the Term 'Chinese Virus'.

[78] During the COVID-19 outbreak, no matter how biased investors were against Asian founders, the worst possible situation was that investors did not open the pitch emails sent by Asian last names and hence the staying time is 0 seconds. This truncation issue at the 0 second will bias our results towards zero. Therefore, it is important to compare magnitudes of race effect when the regression only focuses on opened emails.

last name raises the staying time by 0.38s in the full sample and 2.5s among the opened emails in 04/2020, which accounts for a 10% increase in the staying time. However, the significantly negative coefficients of the interaction term between "Asian Founder" and "March =1" indicate that using an Asian last name reduces the staying time by 0.28s in the full sample and 3s among the opened emails in 04/2020. This accounts for a 12.5% decrease in the staying time. Columns (2) and (3) also suggest that there is a bias against Asian founders in 03/2020 and the direction of bias flipped after 04/2020. Hence, the direction of bias also depends on timing and can be temporarily affected by big societal events, such as the COVID-19 outbreak.

It should be noted that the temporary bias against Asian founders in 03/2020 should be interpreted as the lower bound of the bias in the real world. As shown in Section 3.3, investors are implicitly more biased against Asian founders when stakes are higher. If bias against Asian founders even exists in the pitch email setting, this racial bias can be even worse in other mainstream fundraising settings. Moreover, the experimental setting is very noisy as many factors can affect investors' email behaviors. Hence, the racial bias should be salient enough in order to generate significant results. Both reasons suggest an even harsher fundraising environment for Asians during the COVID-19 outbreak.

4.3 Mechanisms (Testing the Nature of Discrimination)

The mainstream discrimination theories in the correspondence test experiment can be classified into the following three types (Bohren, Imas and Rosenberg (2019b)): belief-based mechanisms, taste-based mechanisms, and amplifying mechanisms. Although results show that the gender bias is likely driven by taste and that the racial bias is mainly driven by belief, multiple subtle mechanisms can coexist. A summary of the related theory predictions and testing results are provided in Appendix C (see Table C9 for gender bias and Table C10 for racial bias).

4.3.1 Belief-Based Mechanisms

Expected Quality and Financial Returns (First Moment) Investors' bias towards female and Asian founders can stem from them foreseeing higher future returns in these founders' startups compared to other founders who also send cold call emails.[79] Table 11 tests this channel and finds that only the bias towards Asians is mainly driven by this belief. Column (4) shows that mentioning an Ivy League educational background reduces the bias towards Asian founders compared to white founders by 0.7% in email opening rates. Column (5) shows that the interaction effect

[79] Investors may hold this belief because of previously documented facts (Ewens and Townsend (2020)), Fairlie and Robb (2010)), the self-selection effect of minority founders (Rosette and Tost (2010), Fernandez and Fogli (2009), Buttner and Moore (1997), Puri and Robinson (2013), Baron, Markman and Hirsa (2001), Fryer Jr (2007), Bohren et al. (2019a), Howell and Nanda (2019), and Kacperczyk and Younkin (2019).), the lower negotiation power of minority founders (Amatucci and Sohl (2004)), more pleasant collaboration experiences (Shane, Dolmans, Jankowski, Reymen and Romme (2012)), etc.

of using an Asian name and mentioning an Ivy League educational background increases to -3.2% if I focus on the sub-sample of emails sent after 03/23/2020 that only mention "pure" Ivy League colleges.[80] This supports the belief-driven bias hypothesis because more signals about the startup's quality will correct this belief and reduce the bias. However, I do not find any suggestive evidence supporting belief-driven gender bias. According to Columns (1)-(3), the interaction term of being a female founder and attending Ivy League colleges is insignificant and even slightly positive.[81]

Expected Variance of Different Groups (Second Moment) According to Heckman and Siegelman [HS] Critique (Siegelman and Heckman (1993); Heckman (1998)), even in the ideal case in which both observed and unobserved group averages (i.e., first moment statistics) are identical, the correspondence test can generate spurious evidence of discrimination in either direction when the belief of unobserved productivity variance differs.[82] Neumark (2012) develops a model that can address this concern and recover an unbiased estimate of discrimination.[83] Table C6 shows that results are still robust after correcting for the source of bias from unobserved variance using Neumark's model. Column (1) demonstrates that using a female name significantly increases the email opening rate by 1%, and I cannot reject the hypothesis that the variances between female and male founders are the same. However, the relative variance of female founders and male founders is smaller than 1, indicating that investors expect female founders to be more homogeneous. Columns (2) and (3) show that using an Asian last name still increases the email opening rate by 0.7% and that the relative variance of Asian founders and white founders decreases from 1.12 in 03/2020 to 1.09 in 04/2020. This means that investors expect Asian-led startups to have more uncertainties than white-led startups during the COVID-19 outbreak. Fortunately, these uncertainties decrease starting in April. In a nutshell, the expected variance of different groups is not the main driver of the detected bias towards minority founders.

Strategic Channel The entrepreneurial financing process in the VC industry is a two-sided matching process (Sørensen (2007)). Theoretically speaking, investors may prefer minority founders if similar majority founders are "over-qualified" and have weaker willingness to collaborate with them due to many outside options.[84] However, Table 10 Columns (4) and (5) rule out this strategic channel by showing that mentioning an excellent educational background still significantly increases investors' email opening rates. In this experimental setting, investors do not reject a startup

[80] President Trump stopped using the phrase "Chinese Virus" on 03/23/2020.

[81] Belief-driven gender bias can still exist if educational background does not explain the performance gap between women-led startups and male-led startups. For example, the recent affirmative actions potentially make the founder's racial information more informative of the startup's quality than the founder's gender information. In this situation, Experiment B cannot detect the belief-driven gender bias.

[82] This is because a standard correspondence test only observes a nonlinear binary decision outcome (i.e., reply vs. no reply, etc.) and this outcome can be affected by higher moment statistics. For example, in a correspondence test design including only high-quality pitch emails, female founders can still receive more replies if investors expect female-led startups to be more homogeneous than male-led startups even if their expected quality is the same between female-led startups and male-led startups.

[83] This model uses a Heteroscedastic Probit Model after imposing several parametric assumptions. I extend his model a little bit by adjusting his assumed monotonic hiring rules. The full discussion and review of this model are provided in Appendix D.

[84] In the labor market, beliefs about the likelihood that candidates will accept job offers constitute a typical confounding mechanism, and employers can reject "over-qualified" candidates due to this strategic channel.

team because the founders are "too good" or "overqualified".

4.3.2 Taste-Based Mechanisms

Friendly Support Investors are likely to be biased towards minority founders because they want to support disadvantaged groups. For example, some impact funds or angel groups, such as 37 Angels, only invest in female-led startups. Table C7 supports this hypothesis and finds that investors working in not-for-profit impact funds are more likely to be biased towards female founders and weakly biased towards Asian founders.[85] Columns (1)-(3) document that the bias towards women is higher for impact funds compared to common funds. Using female names increases the email opening rate by 10.3% for impact fund investors and only 1.1% for common fund investors who do not have special ESG goals. The magnitude of this gender effect for impact funds is roughly 10 times that for common funds. Similarly, Columns (4)-(6) also find that impact funds open more emails sent with an Asian last name, although the magnitude of this racial effect for impact funds is only 2 times the effect for common funds, and results are not very significant.[86] Results above show that the bias towards female and Asian founders partially stems from friendly support from impact funds.

Homophily Homophily means that people prefer groups that share similar backgrounds to themselves (Egan, Matvos and Seru (2017)). Table 12 finds that although male investors are slightly more likely to open pitch emails sent by female names, it is female investors who spend more time on pitch emails sent by female names.[87] Columns (2) and (3) show that using female names increases the email opening rate by 0.8% among female investors and 1.1% among male investors, although Column (1) shows that the difference in female and male investors' responses is not significantly different from zero. Column (4) shows that female investors spend 0.5s more time reading pitch emails sent by a female name. Comparison of Columns (5) and (6) shows that the bias towards female founders as measured by email staying time is four times larger for female investors compared with male investors. However, most results are not significant in this noisy experimental setting. Experiment B also cannot identify other taste-based mechanisms such as the social image effect or the potential sexual harassment concern.[88]

[85] Not-for-profit impact funds are defined using the primary investor type from Pitchbook.

[86] Table C8 shows that this heterogeneous effect between impact funds and common funds still exists for the "keyword" classification method of impact funds. Besides the funds whose primary investor type is "not-for-profit", I also include funds whose stated investment preferences contain keywords such as "ESG," "impact," and "MWBE".

[87] I do not test the homophily effect based on race because the racial information of investors is not provided in the data, and any race prediction algorithms based on names is very noisy.

[88] "Female Founders Still Face Sexual Harassment from Investors," October 15, 2018, shows that among respondents to the survey sent by Y Combinator, more than 20% of women said they had been harassed.

4.3.3 Amplifying Mechanisms (Attention Discrimination, Implicit Bias)

Mechanisms that can magnify both taste-based bias and belief-based bias may also exist. These amplifying mechanisms include attention discrimination and implicit bias. "Attention discrimination" theory (Bartoš et al. (2016)) predicts that even if complete information about an individual is readily available, discrimination can still occur because investors may endogenously allocate their scarce attention to their preferred groups before they make their decisions. Considering that all the outcome variables (i.e., email opening rates, time spent on each pitch email) actually measure the attention spent by investors rather than finalized investment decisions, results naturally support the existence of the attention discrimination channel. "Implicit bias" refers to the attitudes or stereotypes that affect investors' decisions in an unconscious manner. Unfortunately, Experiment B does not provide direct evidence to test this channel.

4.3.4 Alternative Mechanisms

Uninformative Email Replies Investors may pretend to behave more friendly to minorities in the email setting. Hence, these email replies are not indicative of their true investment preferences. However, this hypothesis cannot explain the results found through measuring the email opening rates and email reading time because these behaviors are usually not observed by the founders directly or used in previous correspondence tests. Therefore, I can rule out this mechanism confidently.

4.4 Discussion

Experiment B has the following advantages due to the new email tracking technology. First, it improves the internal validity compared with the standard correspondence test design. By tracking multiple investors' detailed information acquisition behaviors, which were unobservable before, researchers can mitigate the "low-response-rate" problem and introduce meaningful variation in startups' quality. This helps to identify the nature of bias and also to test discrimination during the economic recession period. Second, compared with Experiment A, Experiment B has a relatively stronger external validity since it can recruit a much larger number of investors.

However, Experiment B also suffers from several standard limitations of the correspondence test method. For example, the experimental setting is relatively noisy (Bernstein et al. (2017)). The results from the pitch email setting may not be generalizable to other mainstream fundraising settings as the incentive to elicit investors' preferences is relatively weak. Also, researchers only observe the initial contact stage rather than any later investment stage.[89] After the data

[89]Based on the attention discrimination theory by Bartoš et al. (2016), investors benefit more from spending their scarce attention on their preferred startup groups in a cherry-picking market (i.e., venture capital investment setting). Hence, if no further bias exists in the

confidentiality rule was implemented in the European Union starting in 2018, the correspondence test method, which does not provide a consent form to participants, is infeasible in these areas.

5 Discussion

5.1 Experiment A and Experiment B Are Complementary

Experiment A, as a lab-in-the-field experiment, and Experiment B, as a correspondence test, complement each other in the following sense. First, Experiment A has stronger internal validity (i.e., provides stronger incentives to elicit preferences and identifies the nature of detected biases) and slightly weaker external validity (i.e., recruits a small number of subjects). However, Experiment B has stronger external validity (i.e., recruits a large number of subjects) and weaker internal validity (i.e., struggles to test mechanisms, provides weaker incentives to elicit preferences). Second, Experiment A captures how investors evaluate startups when working with incubators, while Experiment B captures how investors evaluate different cold call pitch emails. Hence, comparing results from both experiments can help researchers to better understand the magnitude of biases in different situations.[90]

5.2 Policy Implications

Results from these experiments provide the following policy implications for handling discrimination problems in the VC industry. First, any training or actions that mitigates *implicit* bias are crucial, as belief-driven implicit bias is the main driver of bias against minority founders. Thus, in fundraising activities like Startup Pitch Night, it is helpful to provide minority founders with earlier time slots for their presentations so investors are not tired yet. Second, the entrepreneurial community and policymakers need to take actions to prevent more serious discrimination problems when the state of the economy worsens. During recessions, VC funds using the relative investment strategy will increase the bar for investment and focus on the most high-quality startups. However, biases against minority founders mainly exist when investors evaluate these attractive startups. Third, emphasizing the success of female/Asian founders and any type of training that improves minority founders' persuasion skills are helpful to mitigate belief-driven discrimination. Fourth, investors should be encouraged to invite more minority founders to enter the communication stage and learn more details of these startups.[91] Lastly, increasing impact funds and the diversity of the VC industry helps.

later-round communication stages, the amount of attention measured is indicative of investors' internal preferences. Some research work (Hu and Ma (2020), Kanze, Huang, Conley and Higgins (2018)) analyzes video data to study the later communication stage.

[90] Generally, the lab-in-the-field experiment also has a more ethical design at the cost of suffering from the potential consent form effect, while the correspondence test involves deception to avoid the potential consent form effect. Future researchers can make an effort to improve internal validity and external validity of both experiments.

[91] Special thanks to Paul Beaumont for pointing this out. Since most experimental results in this paper focus on the initial contact stage rather than the investment stage, it indicates that minority founders have fewer opportunities to elaborate on details of their startups

6 Conclusion

This paper studies whether early-stage investors are biased against female, Asian, and older founders during the investment process. Despite the importance of this question, the literature has mixed results, and there is scarce causal evidence. To identify bias and its nature in the venture capital industry, this paper implements two complementary field experiments with real U.S. venture capitalists. The first experiment (i.e., Experiment A) invites investors to evaluate multiple startup profiles, which they know to be hypothetical, in order to be matched with appropriate startups from collaborative incubators. Investors can also use the tool to donate a small amount of money to randomly displayed startup teams. To test the external validity, the second experiment (i.e., Experiment B) recruits a larger number of investors. It exploits new email behavior tracking technologies to compare investors' detailed information acquisition behaviors when evaluating cold call pitch emails with randomized startups' information.

Results prove the existence of biases based on the founders' group membership and identify its nature. Moreover, the findings also provide one explanation to reconcile the contradictory results in the literature by demonstrating how the biases vary in different settings. Basically, I find that investors have implicit biases *against* female and Asian founders, especially when they evaluate relatively high-quality startups. However, investors can be biased *towards* these minority founders if their startups are relatively low-quality. Hence, previous literature may only capture a specific part of a larger picture, leading to mixed results. Second, among multiple coexisting mechanisms, statistical discrimination is an important reason for "anti-minority" investors' decisions. The paper also finds a taste-driven gender homophily effect as both male and female investors are more likely to provide anonymous support to founders of their same gender. Lastly, the paper detects a temporary, stronger bias against Asian founders during the COVID-19 outbreak, which started to fade after April 2020. This indicates that big societal events can have a temporary impact on the direction of group-level biases.

Overall, the paper contributes to the debate about discrimination in the venture capital industry and provides an experimental framework to detect discrimination and its nature in a financial market setting. Future researchers can test whether such bias also exists in other parts of the entrepreneurial financing system and investigate its implication for equilibrium outcomes. Also, studies exploiting any "natural experiment" settings are helpful to implement welfare analysis.

---

because they are less likely to enter the communication stage. This is a form of "attention discrimination" as investors collect less information about minority founders' companies.

References

Altonji, Joseph G and Rebecca M Blank, "Race and gender in the labor market," *Handbook of labor economics*, 1999, *3*, 3143–3259.

Amatucci, Frances M and Jeffrey E Sohl, "Women entrepreneurs securing business angel financing: Tales from the field," *Venture Capital*, 2004, *6* (2-3), 181–196.

Armona, Luis, Andreas Fuster, and Basit Zafar, "Home price expectations and behaviour: Evidence from a randomized information experiment," *The Review of Economic Studies*, 2019, *86* (4), 1371–1410.

Arrow, Kenneth et al., "The theory of discrimination," *Discrimination in labor markets*, 1973, *3* (10), 3–33.

Barber, Brad M, Adair Morse, and Ayako Yasuda, "Impact investing," *Journal of Financial Economics*, 2020.

____ **, Wei Jiang, Adair Morse, Manju Puri, Heather Tookes, and Ingrid M Werner**, "What Explains Differences in Finance Research Productivity During the Pandemic?," *The Journal of Finance*, 2021.

Baron, Robert A, Gideon D Markman, and Azita Hirsa, "Perceptions of women and men as entrepreneurs: evidence for differential effects of attributional augmenting.," *Journal of Applied psychology*, 2001, *86* (5), 923.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka, "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, June 2016, *106* (6), 1437–1475.

Becker, Gary S, "Nobel lecture: The economic way of looking at behavior," *Journal of political economy*, 1993, *101* (3), 385–409.

____ , *The economics of discrimination*, University of Chicago press, 2010.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws, "Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment," *The Journal of Finance*, 2017, *72* (2), 509–538. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12470.

____ **, Xavier Giroud, and Richard R Townsend**, "The impact of venture capital monitoring," *The Journal of Finance*, 2016, *71* (4), 1591–1622.

Bertrand, Marianne, "Gender in the Twenty-First Century," in "AEA Papers and Proceedings," Vol. 110 2020, pp. 1–24.

____ **and Esther Duflo**, "Field experiments on discrimination," in "Handbook of economic field experiments," Vol. 1, Elsevier, 2017, pp. 309–393.

____ **and Kevin F Hallock**, "The gender gap in top corporate jobs," *ILR Review*, 2001, *55* (1), 3–21.

____ **and Sendhil Mullainathan**, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American economic review*, 2004, *94* (4), 991–1013.

____ **, Claudia Goldin, and Lawrence F Katz**, "Dynamics of the gender gap for young professionals in the financial and corporate sectors," *American economic journal: applied economics*, 2010, *2* (3), 228–55.

——, **Dolly Chugh, and Sendhil Mullainathan**, "Implicit discrimination," *American Economic Review*, 2005, *95* (2), 94–98.

——, ——, **and** ——, "Implicit Discrimination," *American Economic Review*, April 2005, *95* (2), 94–98.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, 1992, *100* (5), 992–1026.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg, "The dynamics of discrimination: Theory and evidence," *American economic review*, 2019, *109* (10), 3395–3436.

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

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

——, ——, ——, **and** ——, "Beliefs about gender," *American Economic Review*, 2019, *109* (3), 739–73.

Brock, Michelle and Ralph De Haas, "Discriminatory Lending: Evidence from Bankers in the Lab," 2020.

Buttner, E Holly and Dorothy P Moore, "Women's organizational exodus to entrepreneurship: self-reported motivations and correlates with success," *Journal of small business management*, 1997, *35*, 34–46.

Carpenter, Jeffrey, Cristina Connolly, and Caitlin Knowles Myers, "Altruistic behavior in a representative dictator experiment," *Experimental Economics*, 2008, *11* (3), 282–298.

Carrell, Scott E, Marianne E Page, and James E West, "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Casaburi, Lorenzo and Jack Willis, "Time versus state in insurance: Experimental evidence from contract farming in Kenya," *American Economic Review*, 2018, *108* (12), 3778–3813.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter, "Race and economic opportunity in the United States: An intergenerational perspective," *The Quarterly Journal of Economics*, 2020, *135* (2), 711–783.

Cunningham, Tom and Jonathan de Quidt, "Implicit preferences inferred from choice," *Available at SSRN 2709914*, 2015.

DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao, "The importance of being marginal: Gender differences in generosity," *American Economic Review*, 2013, *103* (3), 586–90.

Egan, Mark L, Gregor Matvos, and Amit Seru, "When Harry fired Sally: The double standard in punishing misconduct," Technical Report, National Bureau of Economic Research 2017.

Ewens, Michael and Richard R. Townsend, "Are early stage investors biased against women?," *Journal of Financial Economics*, March 2020, *135* (3), 653–677.

Fairlie, Robert W and Alicia M Robb, *Race and entrepreneurial success: Black-, Asian-, and White-owned businesses in the United States*, MIT Press, 2010.

Fang, Hanming and Andrea Moro, "Theories of statistical discrimination and affirmative action: A survey," in "Handbook of social economics," Vol. 1, Elsevier, 2011, pp. 133–200.

Fernandez, Raquel and Alessandra Fogli, "Culture: An empirical investigation of beliefs, work, and fertility," *American economic journal: Macroeconomics*, 2009, *1* (1), 146–77.

Floyd, Eric and John A List, "Using field experiments in accounting and finance," *Journal of Accounting Research*, 2016, *54* (2), 437–475.

Giuliano, Laura, David I Levine, and Jonathan Leonard, "Manager race and the race of new hires," *Journal of Labor Economics*, 2009, *27* (4), 589–631.

Gneezy, Uri, John List, and Michael K Price, "Toward an understanding of why people discriminate: Evidence from a series of natural field experiments," Technical Report, National Bureau of Economic Research 2012.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

Gompers, Paul A and Sophie Q Wang, "Diversity in Innovation," Working Paper 23082, National Bureau of Economic Research January 2017. Series: Working Paper Series.

____ and ____ , "Diversity in innovation," Technical Report, National Bureau of Economic Research 2017.

____ , Vladimir Mukharlyamov, Emily Weisburst, and Yuhai Xuan, "Gender effects in venture capital," *Available at SSRN 2445497*, 2014.

Gornall, Will and Ilya A. Strebulaev, "Gender, Race, and Entrepreneurship: A Randomized Field Experiment on Venture Capitalists and Angels," SSRN Scholarly Paper ID 3301982, Social Science Research Network, Rochester, NY March 2020.

____ and Ilya A Strebulaev, "Squaring venture capital valuations with reality," *Journal of Financial Economics*, 2020, *135* (1), 120–143.

Guzman, Jorge and Aleksandra Olenka Kacperczyk, "Gender gap in entrepreneurship," *Research Policy*, 2019, *48* (7), 1666–1680.

Hebert, Camille, "Gender stereotypes and entrepreneur financing," in "10th Miami Behavioral Finance Conference" 2020.

Heckman, James J, "Detecting discrimination," *Journal of economic perspectives*, 1998, *12* (2), 101–116.

Hegde, Deepak, Alexander Ljungqvist, and Manav Raj, "Race, Glass Ceilings, and Lower Pay for Equal Work," *Available at SSRN 3811410*, 2021.

Henderson, Loren, Cedric Herring, Hayward Derrick Horton, and Melvin Thomas, "Credit where credit is due?: Race, gender, and discrimination in the credit scores of business startups," *The Review of Black Political*

Economy, 2015, *42* (4), 459–479.

Hong, Harrison and Inessa Liskovich, "Crime, punishment and the halo effect of corporate social responsibility," Technical Report, National Bureau of Economic Research 2015.

Howell, Sabrina T. and Ramana Nanda, "Networking Frictions in Venture Capital, and the Gender Gap in Entrepreneurship," SSRN Scholarly Paper ID 3376211, Social Science Research Network, Rochester, NY October 2019.

____, **Josh Lerner, Ramana Nanda, and Richard Townsend**, "Financial Distancing: How Venture Capital Follows the Economy Down and Curtails Innovation," SSRN Scholarly Paper ID 3594239, Social Science Research Network, Rochester, NY May 2020.

Hu, Allen and Song Ma, "Human Interactions and Financial Investment: A Video-Based Approach," *Available at SSRN*, 2020.

Jr, Roland G Fryer, "Belief flipping in a dynamic model of statistical discrimination," *Journal of Public Economics*, 2007, *91* (5-6), 1151–1166.

____ **and Steven D Levitt**, "The causes and consequences of distinctively black names," *The Quarterly Journal of Economics*, 2004, *119* (3), 767–805.

Kacperczyk, Aleksandra and Peter Younkin, "The Illegitimacy Premium: The Effect of Entrepreneurship on the Future Employment of Women," *Available at SSRN 3433249*, 2019.

Kanze, Dana, Laura Huang, Mark A Conley, and E Tory Higgins, "We ask men to win and women not to lose: Closing the gender gap in startup funding," *Academy of Management Journal*, 2018, *61* (2), 586–614.

Kessler, Judd B., Corinne Low, and Colin D. Sullivan, "Incentivized Resume Rating: Eliciting Employer Preferences without Deception," *American Economic Review*, November 2019, *109* (11), 3713–3744.

Lang, Kevin and Ariella Kahn-Lang Spitzer, "Race discrimination: An economic perspective," *Journal of Economic Perspectives*, 2020, *34* (2), 68–89.

____ **and Jee-Yeon K Lehmann**, "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Literature*, 2012, *50* (4), 959–1006.

Lee, Sokbae and Bernard Salanié, "Identifying effects of multivalued treatments," *Econometrica*, 2018, *86* (6), 1939–1963.

List, John A, "The nature and extent of discrimination in the marketplace: Evidence from the field," *The Quarterly Journal of Economics*, 2004, *119* (1), 49–89.

Lundberg, Shelly J and Richard Startz, "Private discrimination and social intervention in competitive labor market," *The American Economic Review*, 1983, *73* (3), 340–347.

Neumark, David, "Detecting Discrimination in Audit and Correspondence Studies," *The Journal of Human Resources*, 2012, *47* (4), 1128–1157. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System].

Phelps, Edmund S, "The statistical theory of racism and sexism," *The american economic review*, 1972, *62* (4), 659–661.

Puri, Manju and David T Robinson, "The economic psychology of entrepreneurship and family business," *Journal of Economics & Management Strategy*, 2013, *22* (2), 423–444.

Raina, Sahil, "VCs, founders, and the performance gender gap," *Finance Down Under 2017 Building on the Best from the Cellars of Finance*, 2019.

Rosette, Ashleigh Shelby and Leigh Plunkett Tost, "Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders.," *Journal of Applied Psychology*, 2010, *95* (2), 221.

Shane, Scott, Sharon Dolmans, Joseph Jankowski, Isabelle Reymen, and Georges Romme, "Which inventors do technology licensing officers favor for start-ups," *Frontiers of Entrepreneurship Research*, 2012, *32* (18), 1.

Siegelman, Peter and J Heckman, "The Urban Institute audit studies: Their methods and findings," *Clear and Convincing Evidence: Measurement of Discrimination in America, Washington*, 1993, *187*, 258.

Sørensen, Morten, "How smart is smart money? A two-sided matching model of venture capital," *The Journal of Finance*, 2007, *62* (6), 2725–2762.

Zhang, Ye, "How Venture Capitalists Bet: Evidence from Two Randomized Controlled Trials," *Available at SSRN 3724424*, 2020.

____, "Impact Investing and Venture Capital Industry: Experimental Evidence," *Available at SSRN 3959117*, 2021.

Tables

Table 1: Summary Statistics for Investors

Panel A: Investor Location Distribution

Country	N	Percentage	Female Percentage
US	15,184	84.91%	23.57%
Canada	647	3.62%	29.68%
Israel	456	2.55%	29.39%
UK	93	0.52%	22.58%
India	514	2.87%	18.87 %
Singapore & Hong Kong	454	2.54%	21.59%
Australia & New Zealand	228	1.28%	25.44%
Others	306	1.71%	21.57%
Total	17882	100%	

Panel B: Investor Industry Distribution

Industry	N	Percentage
Information Technology	13,628	76.21%
Healthcare	6,056	33.87%
Consumers	6,256	34.98%
Energy	4,234	23.68%
Life Sciences	3,347	18.72%
Finance	3,023	16.91%
Media & Entertainment	2,533	14.17%
Agriculture & Food	2,072	11.59%
Transportation	1,743	9.75%
Education	1,359	7.60%
Clean Technology	1,201	6.72%
Others	3,271	18.29%

Panel C: Investor Characteristics

	N	Mean
Female Investor=1	17,882	0.24
Senior Investor=1	17,882	0.84
Angel Investor=1	17,882	0.11
Top University=1	13,785	0.31
Graduate School=1	9,232	0.61
Not-for-profit Fund=1	13,156	0.02

Notes. This table reports descriptive statistics for active venture capitalists (defined as those whose email addresses are verified by the testing email) who received cold call pitch emails in the correspondence test and recruitment emails in the lab-in-the-field experiment. Panel A reports the geographical distribution of the sample investors. "Others" includes South Africa, Cayman Islands, Malaysia, etc. Panel B reports the industries that recruited investors are interested in. An investor can indicate multiple preferred industries. "Others" includes special industries, such as packaging technology industry. 3.8% of investors' industry preferences cannot be found online and I have assumed that they are interested in all the industries when sending out cold call pitch emails. Panel C reports investors' demographic information and investment philosophies. "Female = 1" is an indicator variable that equals one if the investor is female, and zero otherwise. "Senior = 1" is an indicator variable that equals one if the investor is senior (defined as C-level positions, principals, vice president, partners, etc.), and zero otherwise. "Angel = 1" is an indicator variable that equals one if the investor is an angel investor or belongs to an angel group, and zero otherwise. If an investor is both an angel investor and also an institutional investor, I treat her as an angel investor. "Not-for-profit Fund = 1" is an indicator variable that equals one if the investor works in a not-for-profit impact fund based on the "primary investor type" in the Pitchbook Database. "Top University = 1" and "Graduate School = 1" are indicator variables that equal one if the investor has attended a top university (i.e., Ivy League colleges, MIT, Duke, Caltech, Amherst, Northwestern, Stanford, UC Berkeley, University of Chicago and Williams College) or has attended graduate school.

Table 2: Experiment A Summary Statistics of Investors

Panel A: Investor Stated Interest Across Sectors

Sector (Repeatable)	N	Fraction (%)
Information technology	39	55.7%
Consumers	10	14.3%
Healthcare	17	24.3%
Clean technology	3	4.3%
Business-to-business	7	10.0%
Finance	11	15.7%
Media	4	5.8%
Energy	5	7.1%
Education	3	4.3%
Life sciences	2	2.9%
Transportation & Logistics	4	5.7%
Others	6	8.6%
Industry Agnostic	6	8.6%

Panel B: Investor Stated Interest Across Stages

Stage (Repeatable)	N	Fraction (%)
Seed Stage	47	67.1%
Series A	45	64.3%
Series B	17	24.3%
Series C or later stages	5	7.1%

Panel C: Investor Stated Demographic Information

	N	Mean	S.D
Female Investor	69	0.20	0.40
Minority Investor	64	0.42	0.50
Senior Investor	69	0.86	0.37

Panel D: Investor Investment Philosophy

	N	Mean	S.D
Cold Email Acceptance	69	0.74	0.44
Prefer ESG	69	0.17	0.14
Direct Investment	69	0.94	0.24

Continued

Panel E: Available Fund's Financial Performance

	N	Mean	S.D	Percentile 10	Percentile 50	Percentile 90
Total Active Portfolio	54	41.40	44.51	10	24	102
Total Exits	46	32.74	48.39	1	9	110
Fund Age	52	11.75	8.95	3	8.5	25
AUM (Unit: $1 Million)	33	547.46	1029.10	30	111.7	1700
Dry Power (Unit: $1 Million)	33	163.86	307.04	6.43	44.35	313.59

*Notes.*This table reports descriptive statistics for the investors who have participated in the lab-in-the-field experiment (i.e., Experiment A). In total, 69 different investors from 68 institutions, mostly venture funds, provided evaluations of 1216 randomly generated startup profiles. Panel A reports the sector distribution of investors. Each investor can indicate their interest in multiple industries. "Others" includes HR tech, Property tech, infrastructure, etc. "Industry Agnostic" means the investor does not have strong preferences based on sector. Panel B reports the stage distribution of investors, and each investor can invest in multiple stages. "Seed Stage" includes pre-seed, angel investment, and late-seed stages. "Series C or later stages" includes growth capital, series C, D, etc. Panel C reports the demographic information of these recruited investors. "Female" is an indicator variable which equals to one if the investor is female, and zero otherwise. "Minority" is an indicator variable which equals to one if the investor is Asian, Hispanic, or African Americans, and zero otherwise. Investors who prefer not to disclose their gender or race are not included in these variables. "Senior" is equal to one if the investor is in a C-level position, or is a director, partner, or vice president. It is zero if the investor is an analyst (intern) or associate investor. "Cold Email Acceptance" is an indicator variable which equals one if the investor feels that sending cold call emails is acceptable as long as they are well-written, and zero if the investor feels that it depends. "Prefer ESG" is an indicator variable which equals one if the investor prefers ESG related startups, and zero otherwise. "Direct Investment" is an indicator variable which equals to one if the investor can directly make the investment, and zero if their investment is through limited partners or other channels. Panel E provides the financial information of the 68 funds that these investors work for. However, we can only recover parts of their financial information from the Pitchbook Database.

Table 3: Experiment A Design, Randomization of Profile Components

Profile Component	Randomization Description	Analysis Variable
Startup Team Characteristics		
First and Last Names	Drawn from list of the same names given selected race and gender as used in Experiment 1 (See names in Tables B1)	White Female[a] (25%) Asian Female (25%) White Male (25%) Asian Male (25%)
Number of Founders	The team can have 1 founder or 2 co-founders	Single Founder (8/16)
Age	Founders' age is indicated by the graduation year Young VS Old=50% VS 50% Young: uniformly distributed (2005-2019) Old: uniformly distributed (1980-2005)	Age
Educational Background	Drawn from top school list and common school list (See school list Table B2)	Top School (8/16)
Entrepreneurial Experiences	The team can have serial founder(s) or only first-time founder(s)	Serial Founder (8/16)
Startup Project Characteristics		
Company Age	Founding dates are randomly drawn from the following four years {2016, 2017, 2018, 2019}	Company Age
Comparative Advantages	Randomly drawn from a comparative advantage list (See Tables B3), the number of drawn advantages is between 1 to 4	1 Advantages (4/16) 2 Advantages (4/16) 3 Advantages (4/16) 4 Advantages (4/16)
Traction	Half randomly selected profiles generate no revenue Half randomly selected profiles generate positive revenue. Previous monthly return: uniform distribution [5K, 80K]; Growth rate: uniform distribution [5%, 60%]	Positive traction (8/16)
Company Category	Randomly assigned as either B2B or B2C	B2B (8/16)
Number of Employees	Randomly assigned with one of four categories	0-10 (8/16) 10-20 (8/16) 20-50 (8/16) 50+ (8/16)
Target Market	Randomly assigned as either domestic market or international market	Domestic (8/16)
Mission	Randomly assigned with one of three categories "For profit", "For profit, consider IPO within 5 years", "Besides financial gains, also cares ESG"	For profit (8/16) For profit, IPO Plan (4/16) For profit, ESG (4/16)
Location	Randomly assigned as either U.S. or Outside the U.S.	U.S. (70%)
Previous Funding Situation		
Number of Existing Investors	Randomly assigned as one of the four categories with equal probability {0,1,2,3+}	Number of investors

[a]The randomization distribution is to increase the experimental power. Considering that our collaborating incubators have more Asian and female founders than the normal gender and race distribution, I increased the ratio of female and Asian founders in this experiment to mimic the distribution of these incubators, which provides the pool of potential matched startups. Although some investors feel that providing more information would be helpful, no one complains that the distribution of founding team gender and race is unrealistic.[b]If there are two co-founders in the same founding team, all the founders' background information is similar to each other. For example, if the first founder's age belongs to the young founder category, then the second founder's age also belongs to the same age category.

Notes. This table provides the randomization of each startup profile's components and its corresponding analysis variables. Profile components are listed in the order that they appear on the hypothetical startup profiles. Weights of characteristics are shown as fractions when they are fixed across subjects (e.g., each subject should see exactly 8/16 resumes with all female team members) and percentages when they represent a draw from a probability distribution (e.g., for startups with positive revenue records, the revenue follows a uniform distribution between [5K - 80 K]). Variables in the right-hand column are randomized to test how investors respond to these analysis variables.

Table 4: Experiment A Evaluation Results About Gender, Race and Age

Dependent Variable	Q1 Quality (1)	Q2 Availability (2)	Q3 Contact (3)	Q4 Investment (4)	Q5 Risk (5)
Panel A: Gender					
Female Founder	-0.56	0.46	-0.94	0.04	3.37
	(1.20)	(0.89)	(1.41)	(0.21)	(3.07)
Investor FE	Yes	Yes	Yes	Yes	Yes
Control Mean	44.30	63.84	55.00	6.02	65.19
Profile Observations	1,216	1,184	1,216	1,176	176
R-squared	0.31	0.53	0.47	0.34	0.25
Panel B: Race					
Asian Founder	0.05	-0.61	-0.34	-0.04	0.70
	(1.19)	(0.89)	(1.40)	(0.21)	(3.09)
Investor FE	Yes	Yes	Yes	Yes	Yes
Control Mean	44.31	65.51	55.51	6.12	67.14
Profile Observations	1,216	1,184	1,216	1,176	176
R-squared	0.31	0.53	0.47	0.34	0.24
Panel C: Age					
Age	-0.12	-0.24	-0.35	-0.01	-2.39*
	(0.46)	(0.35)	(0.53)	(0.08)	(1.21)
Age^2	0.00	0.00	0.00	0.00	0.03*
	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Investor FE	Yes	Yes	Yes	Yes	Yes
Observations	1,216	1,184	1,216	1,176	176
R-squared	0.31	0.53	0.47	0.34	0.26

Notes. This table describes evaluation results combining the total profile evaluations, including all the profiles in the first half and all the profiles in the second half. Some investors skip the evaluation questions of availability or investment if they feel the information is not enough to make their judgements. Q5 (risk evaluation) is only added to a randomly selected investor for robustness check. Panel A shows investors' attitudes based on founders' gender information. "Female Founder" is a dummy variable that is equal to one if the startup founder has a female first name, and zero otherwise. Panel B shows investors' attitudes based on founder's race information. "Asian Founder" is a dummy variable that is equal to one if the startup founder has an Asian last name, and zero otherwise. Panel C shows investors' attitudes based on founder's age. Age is the approximated founder's age based on the graduation year from the college. Age^2 is the square of founder's age. In Column (1), the dependent variable is the quality evaluation, which indicates the percentile rank of each startup profile compared with an investor's previous invested startups in terms of its potential financial returns. In Column (2), the dependent variable is the availability evaluation, which indicates how likely the investors think the startup team will accept his/her investment rather than other investors. In Column (3), the dependent variable is the contact interest, which describes the probability that the investor wants to contact this startup. In Column (4), the dependent variable is the relative investment interest ranging from 1 to 20, which describes the relative investment amount compared with the investor's general investment amount. The unit is one-tenth of the relative investment compared with investors' average investment amount. For example, if the investor's average invested deal is \$1M and Q4 is equal to 5, then it means the investor only wants to invest \$1M $\times$ 5 $\times$ 10% = \$500,000 in this startup. If Q4 is 20, then the investment amount is \$1M $\times$ 20 $\times$ 10% = \$2M. In Column (5), the dependent variable is the risk evaluation, which describes the percentile rank of each startup profile compared with an investor's previous invested startups in terms of its risk level. All the regressions add the investor fixed effect. Standard errors in parentheses are robust standard errors. Results are still robust when clustering standard errors on the investor level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 5: Experiment A Implicit Bias Based on Founder's Gender, Race and Age

Dependent Variable	Response Time (Unit: Second) (1)	Q1 Quality (2)	Q2 Availability (3)	Q3 Contact (4)	Q4 Investment (5)	Q5 Risk (6)
Panel A: Gender						
Second Half of Study	-27.20***	2.42	2.27*	0.85	0.95***	-2.83
	(2.29)	(1.63)	(1.25)	(1.97)	(0.29)	(4.11)
Female Founder	-1.34	1.56	1.27	0.89	0.56*	2.14
	(2.31)	(1.69)	(1.33)	(2.02)	(0.30)	(4.50)
Female Founder ×		-4.26*	-1.67	-3.67	-1.03**	2.75
Second Half of Study		(2.42)	(1.79)	(2.84)	(0.43)	(6.21)
p-value of Female Founder in the second half of study		0.11	0.74	0.16	0.12	0.25
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,216	1,216	1,184	1,216	1,176	176
R-squared	0.34	0.31	0.53	0.47	0.35	0.25
Panel B: Race						
Second Half of Study	-27.20***	2.37	1.88	-0.28	0.76***	-4.59
	(2.28)	(1.68)	(1.22)	(1.98)	(0.29)	(4.11)
Asian Founder	0.54	2.26	-0.14	0.41	0.31	-3.17
	(2.35)	(1.70)	(1.34)	(2.04)	(0.30)	(4.47)
Asian Founder ×		-4.41*	-0.93	-1.51	-0.69	7.59
Second Half of Study		(2.44)	(1.82)	(2.88)	(0.43)	(6.25)
p-value of Asian Founder in the second half of study		0.21	0.37	0.58	0.21	0.30
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,216	1,216	1,184	1,216	1,176	176
R-squared	0.34	0.31	0.53	0.47	0.35	0.25

Continued

Dependent Variable	Response Time (Unit: Second) (1)	Q1 Quality (2)	Q2 Availability (3)	Q3 Contact (4)	Q4 Investment (5)	Q5 Risk (6)
Panel C: Age						
Second Half of Study	-27.20***	-7.64	-20.43	-8.34	0.21	81.52*
	(2.28)	(18.86)	(14.30)	(22.28)	(3.26)	(48.78)
Age	-0.18	-0.37	-0.83	-0.51	-0.03	-0.23
	(0.85)	(0.70)	(0.54)	(0.82)	(0.12)	(1.64)
Age^2	0.00	0.00	0.01	0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.02)
Age $\times$ Second Half of Study		0.48	1.10	0.30	0.03	-4.23*
		(0.94)	(0.71)	(1.09)	(0.16)	(2.44)
Age^2 $\times$ Second Half of Study		-0.01	-0.01	-0.00	-0.00	0.05*
		(0.01)	(0.01)	(0.01)	(0.00)	(0.03)
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,216	1,216	1,184	1,216	1,176	176
R-squared	0.34	0.31	0.53	0.47	0.35	0.27

Notes. This table reports regression results of how investors' response time and evaluation results respond to a startup founder's gender and race. Panel A tests the implicit bias based on founder's gender. Panel B tests the implicit bias based on founder's race. Panel C tests the implicit bias based on founder's age. "Female Founder" is a dummy variable that is equal to one if the startup founder has a female first name, and zero otherwise. "Asian Founder" is a dummy variable that is equal to one if the startup founder has an Asian last name, and zero otherwise. "Second Half of Study" is an indicator variable for startup profiles shown among the last eight resumes viewed by a subject. "Age" is the approximated founder's age based on the graduation year from the college. "Age^2" is the square of founder's age. Fixed effects for subjects are included in all specifications. In column (1), the dependent variable is investors' response time, which is defined as the number of seconds before each page submission, winsorized at the 95th percentile (59.23 seconds on average). Columns (2)-(6) show the quality evaluation, availability evaluation, contact interest, investment interest and risk evaluation separately. R-squared is indicated for each OLS regression. Standard errors in parentheses are robust standard errors. Results are still robust when clustering standard errors on the investor level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 6: Experiment A Implicit Bias Based on Founder's Gender by Investors' Industry

Dependent Variable	Response Time	Q1	Q2	Q3	Q4	Q5
	(Unit: Second)	Quality	Availability	Contact	Investment	Risk
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Tech Sector Investors						
Second Half of Study	-24.87***	2.96	4.63***	-0.56	1.18***	-9.18*
	(2.81)	(2.11)	(1.70)	(2.63)	(0.38)	(4.72)
Female Founder	1.28	2.73	1.71	-0.16	0.53	9.12**
	(2.82)	(2.16)	(1.78)	(2.62)	(0.38)	(4.51)
Female Founder ×		-6.59**	-3.28	-3.87	-1.21**	1.21
Second Half of Study		(3.16)	(2.47)	(3.83)	(0.56)	(6.71)
p-value of Female Founder in the second half of study		0.09	0.35	0.14	0.10	0.04
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	784	784	752	784	774	112
R-squared	0.31	0.31	0.41	0.41	0.33	0.40
Panel B: Ivy-scaled Coefficient (In the Second Half of Study)						
Ivy League College		8.78***	-0.48	8.65***	1.20***	-10.69**
		(1.67)	(1.12)	(1.96)	(0.31)	(4.17)
Female Founder/Ivy League College		-0.44	3.22	-0.37	-0.60	-0.88
Panel C: Non-tech Sector Investors						
Second Half of Study	-31.58***	1.49	-1.88	3.56	0.51	7.97
	(3.92)	(2.53)	(1.74)	(2.79)	(0.45)	(7.39)
Female Founder	-6.30	-0.41	0.38	3.03	0.59	-6.81
	(3.97)	(2.71)	(1.94)	(3.10)	(0.49)	(8.65)
Female Founder ×		-0.24	1.21	-3.48	-0.69	2.88
Second Half of Study		(3.69)	(2.41)	(4.02)	(0.65)	(11.23)
p-value of Female Founder in the second half of study		0.80	0.25	0.86	0.82	0.60
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	432	432	432	432	402	64
R-squared	0.37	0.30	0.71	0.54	0.38	0.18

Notes. This table reports regression results of how the response time and evaluation results of investors from different industries respond to a startup founder's gender. Panel A tests the implicit bias of investors working in the tech sectors (i.e., IT, cyber security, software, etc.). Panel B calculates the relative magnitude of the implicit bias in tech sectors compared with the effect of going to an Ivy League college by using the profiles in the second half of the study. Panel C tests the implicit bias of investors working in non-tech sectors (i.e., media, entertainment, education, etc.). "Female Founder" is a dummy variable that is equal to one if the startup founder has a female first name, and zero otherwise. "Second Half of Study" is an indicator variable for startup profiles shown among the last eight resumes viewed by a subject. "Ivy League College" is a dummy variable that is equal to one if the startup founder has graduated from an Ivy League college, and zero otherwise. In column (1), the dependent variable is investors' response time, which is defined as the number of seconds before each page submission, winsorized at the 95^{th} percentile (59.23 seconds on average). Columns (2)-(6) show the quality evaluation, availability evaluation, contact interest, investment interest and risk evaluation separately. R-squared is indicated for each OLS regression. Standard errors in parentheses are robust standard errors. Results are still robust when clustering standard errors on the investor level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 7: Experiment A Implicit Bias Based on Founder's Race by Investors' Contact Interest

Dependent Variable	Response Time (Unit: Second) (1)	Q1 Quality (2)	Q2 Availability (3)	Q3 Contact (4)	Q4 Investment (5)	Q5 Risk (6)
Panel A: Contact Interest is High ($Q3 >= 50$)						
Second Half of Study	-28.05***	4.85**	-0.43	1.83	0.90***	-3.63
	(3.00)	(1.88)	(1.16)	(1.41)	(0.33)	(4.60)
Asian Founder	-0.58	3.51*	-1.36	0.92	0.58*	-1.08
	(3.15)	(1.95)	(1.33)	(1.57)	(0.35)	(5.55)
Asian Founder × Second Half of Study		-7.94***	0.02	-3.66*	-1.44***	6.56
		(2.76)	(1.79)	(2.20)	(0.51)	(7.23)
p-value of Asian Founder in the second half of study		0.02	0.25	0.06	0.01	0.25
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	724	724	692	724	698	127
R-squared	0.37	0.41	0.68	0.44	0.45	0.20
Panel B: Ivy-scaled Coefficient (In the Second Half of Study)						
Ivy League College		8.78***	-0.48	8.65***	1.20***	-10.69**
		(1.67)	(1.12)	(1.96)	(0.31)	(4.17)
Asian Founder/Ivy League College		-0.49	1.27	-0.38	-0.80	-0.36
Panel C: Contact Interest is Low ($Q3 < 50$)						
Second Half of Study	-26.46***	1.90	3.62	2.33*	1.07***	-1.73
	(3.91)	(1.86)	(2.20)	(1.39)	(0.27)	(4.00)
Asian Founder	2.11	1.91	1.57	2.73*	0.47	-4.57
	(3.90)	(1.82)	(2.32)	(1.61)	(0.30)	(4.15)
Asian Founder × Second Half of Study		-1.26	-3.13	-2.48	-0.22	8.46
		(2.68)	(3.08)	(2.09)	(0.42)	(6.19)
p-value of Asian Founder in the second half of study		0.72	0.44	0.85	0.39	0.33
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	492	492	492	492	478	49
R-squared	0.33	0.48	0.60	0.50	0.62	0.87

Notes. This table reports regression results of how investors' response time and evaluation results respond to a startup founder's race in the "high contact interest" situations and the "low contact interest" situations. Panel A tests the implicit racial bias in the "high contact interest" situations in which investors' contact interest is higher than or equal to 50% probability. Panel B calculates the relative magnitude of the implicit racial bias in "high contact interest" situations compared with the effect of going to an Ivy League college by using the profiles in the second half of the study. Panel C tests the implicit racial bias in the "low contact interest" situations in which investors' contact interest is lower than 50% probability. Results are similar if I choose other thresholds like 40% or 45%. "Asian Founder" is a dummy variable that is equal to one if the startup founder has an Asian last name, and zero otherwise. "Ivy League College" is a dummy variable that is equal to one if the startup founder graduates from an Ivy League college, and zero otherwise. "Second Half of Study" is an indicator variable for startup profiles shown among the last eight resumes viewed by a subject. In Column (1), the dependent variable is investors' response time, which is defined as the number of seconds before each page submission, winsorized at the 95th percentile (59.23 seconds on average). Columns (2)-(6) show the quality evaluation, availability evaluation, contact interest, investment interest and risk evaluation separately. Results are still robust when clustering standard errors on the investor level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 8: Experiment A Founder's Gender, Race and Age Contact-Based Heterogeneous Effect

Dependent Variable	(1) Quality	(2) Collaboration	(3) Contact	(4) Investment
Panel A: Gender				
$\beta^3 < 0$ *(Not Contact Female)*				
Female Founder	-16.40***	-2.85	-21.81***	-2.61***
	(2.62)	(1.79)	(2.74)	(0.47)
Ratios of Anti-Women	0.42	0.43	0.42	0.41
$\beta^3 > 0$ *(Contact Female)*				
Female Founder	7.93***	1.54	13.69***	1.08**
	(2.01)	(1.32)	(1.79)	(0.34)
Ratios of Pro-Women	0.58	0.57	0.58	0.59
Panel B: Race				
$\beta^3 < 0$ *(Not Contact Asians)*				
Asian Founder	-12.12***	-1.43	-17.60***	-2.01***
	(2.42)	(1.83)	(2.48)	(0.46)
Ratios of Anti-Asian	0.45	0.46	0.45	0.46
$\beta^3 > 0$ *(Contact Asians)*				
Asian Founder	6.34***	-0.78	12.41***	0.95***
	(2.10)	(1.71)	(2.30)	(0.35)
Ratios of Pro-Asian	0.55	0.54	0.55	0.54
Panel C: Age				
$\beta^3 < 0$ *(Not Contact Older Founders)*				
Older Founder	-13.17***	-1.98	-17.23***	-2.03***
	(2.54)	(1.80)	(2.60)	(0.45)
Ratios of Anti-Older	0.38	0.40	0.38	0.38
$\beta^3 > 0$ *(Contact Older Founders)*				
Older Founder	7.83***	2.06	14.47***	1.34***
	(1.96)	(1.32)	(2.01)	(0.38)
Ratios of Pro-Older	0.62	0.60	0.62	0.62
Investor FE	Yes	Yes	Yes	Yes
Observations	608	592	608	591

Notes. This table reports the contact decision-based heterogeneous effect of startup founder's gender, race and age by using the second half of evaluation questions. Panels A, B, and C report the heterogeneous effect of investors who want to contact female, Asian, and older founders and those who want to contact male, white, and younger founders, respectively. "Female Founder" is an indicative variable that is equal to one if the startup founder is female, and zero otherwise. Ratios of "Anti-Women" is the number of profiles with $\beta^3 < 0$ divided by the number of profiles used. Ratios of "Pro-Women" is the number of profiles with $\beta^3 > 0$ divided by the total number of profiles used. "Asian Founder" is an indicative variable that is equal to one if the startup founder is Asian, and zero otherwise. Ratios of Anti-Asian is the number of profiles with $\beta^3 < 0$ divided by the total number of profiles used. Ratios of Pro-Asian is the number of profiles with $\beta^3 > 0$ divided by the number of profiles used. "Older Founder" is an indicative variable that is equal to one if the startup founder graduated from college in 2005 or before, and zero otherwise. Ratios of "Anti-Older" is the number of profiles with $\beta^3 < 0$ divided by the total number of profiles used. Ratios of "Pro-Older" is the number of profiles with $\beta^3 > 0$ divided by the number of profiles used. All the regression results are estimated using the "Leave-one-out" estimator after adding the investor fixed effect. Standard errors in parentheses are bootstrapped for the two-stage calculations. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Experiment A Taste-Based Bias from the Donation Section

	Dependent Variable: Donated Amount (Unit:$)					
	Full Sample			With Donation Decisions		
	(1)	(2)	(3)	(4)	(5)	(6)
Female Founder	0.49	-3.05*		0.64	-2.81*	
	(2.27)	(1.70)		(2.29)	(1.65)	
Asian Founder	4.20**		1.04	4.27**		0.37
	(1.71)		(1.87)	(1.64)		(1.75)
Female Founder× Asian Founder	-4.81			-4.70		
	(3.20)			(3.14)		
Female Founder × Female Investor		7.05			10.31***	
		(4.50)			(3.57)	
Asian Founder × Asian Investor			1.05			3.74
			(3.47)			(3.48)
Female Investor	-4.23**	-7.41***		-1.33	-5.38	
	(2.12)	(2.66)		(2.16)	(3.33)	
Asian Investor	-4.07**		-4.71*	-3.75**		-5.33**
	(1.69)		(2.42)	(1.65)		(2.48)
Constant	11.10***	12.41***	10.71***	11.47***	12.88***	12.00***
	(1.34)	(1.09)	(1.32)	(1.37)	(1.02)	(1.24)
Observations	69	69	70	61	61	62
R-squared	0.18	0.12	0.09	0.14	0.10	0.10

Notes. This table reports regression results from the donation section (i.e., the dictator game in Experiment A), which tests whether there is any taste-driven bias based on a startup founder's gender and race when the donation is anonymously implemented. The dependent variable is the donated amount measured in dollars, ranging from \$0 to \$15. In Columns (1)-(3), I include the investors who did not select a donation amount and treat their behaviors as "donate \$0". In Columns (4)-(6), I exclude the investors who did not select a donation amount. "Female Founder" is an indicative variable which equals to one if the displayed startup founder is female, and zero otherwise. "Asian Founder" is an indicative variable which equals to one if the displayed startup founder is Asian, and zero otherwise. "Female Founder× Asian Founder" is the interaction term of "Female Founder" and "Asian Founder". Similarly, "Female Investor" and "Asian Investor" are indicative variables which are equal to one if the investor is female or Asian. All regressions use robust standard errors reported in parentheses., *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 10: Experiment B Investor Responses to Randomized Emails

Panel A: Response Summary Statistics

	N	Mean	Median	S.D.	Min	Max
Open Rate	3,720	12.03%	0	0.33	0	1
Staying Time (Unit: s)	3,381	24.10	10.33	26.73	0.01	86.63
Click Rate	519	1.68%	0	0.13	0	1
Email Replies	472	1.53%				

Panel B: Email Opening Behaviors

	Dependent Variable: 1(*Opened*)				
	(1)	(2)	(3)	(4)	(5)
	Full	Full	Full	"Pure Ivy"	Full
Female Founder=1	0.010***				0.010***
	(0.004)				(0.004)
Asian Founder=1		0.007*			0.006
		(0.004)			(0.004)
Ivy=1			0.007*	0.012**	0.007*
			(0.004)	(0.005)	(0.004)
Project Advantage=1					0.001
					(0.004)
Asian Founder=1 × March Chinese Virus=1		-0.009			
		(0.010)			
March Chinese Virus=1		-0.040**			
		(0.020)			
US Investor=1	-0.016***	-.016***	-0.016***	-0.023***	-0.016***
	(0.006)	(0.006)	(0.006)	(0.008)	(0.006)
Female Investor=1	-0.019***	-0.020***	-0.019***	-0.017***	-0.019***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
Constant	0.193***	0.194***	0.194***	0.108***	0.186***
	(0.019)	(0.019)	(0.019)	(0.017)	(0.019)
Startup FE	Yes	Yes	Yes	Yes	Yes
Observations	30,909	30,909	30,909	16,578	30,909
Adjusted R-squared	0.005	0.005	0.005	0.006	0.005

Continued

Panel C: Staying Time

	Dependent Variable: Staying Time (Unit: s)		
	(1) Full Sample (Gender)	(2) Full Sample (Race)	(3) Opened Emails (Race)
Female Founder=1	0.12	0.25*	0.31
	(0.19)	(0.13)	(0.88)
Asian Founder=1	0.28	0.38**	2.49*
	(0.13)	(0.19)	(1.34)
Ivy=1	0.11	0.11	-0.12
	(0.13)	(0.13)	(0.88)
Project Advantage=1	0.12	0.12	0.92
	(0.13)	(0.13)	(0.88)
US Investor=1	-0.24	-0.24	1.30
	(0.20)	(0.20)	(1.20)
March=1	1.23	1.68*	6.11
	(0.93)	(0.93)	(4.98)
Female Founder=1 × March=1	0.24		
	(0.26)		
Asian Founder=1 × March=1		-0.66**	-5.48***
		(0.26)	(1.74)
Control	Yes	Yes	Yes
Pitch FE	Yes	Yes	Yes
Observations	30,909	30,909	3,720
Adjusted R-squared	0.002	0.003	0.002

Notes. This table summarizes investors' email responses in the correspondence test and reports regression results of global investors' email opening behaviors in response to randomized pitch emails in Experiment B. Panel A summarizes important investors' information acquisition behaviors in the pitch email setting. Panel B reports regression results of how startup characteristics affect investors' email opening behaviors. Panel C reports regression results of how startup characteristics affect investors' staying time on each pitch email. In Panel B, the dependent variable is a dummy variable, which is one when an investor opens the pitch email, and zero otherwise. "Female Founder = 1" is an indicator variable that equals one if the first name of the email sender is a female name, and zero otherwise. Similarly, "Asian Founder = 1" is an indicator variable that equals one if the last name of the email sender is an East Asian name, and zero otherwise. "Ivy = 1" is an indicator variable for Ivy League educational background. "Project Advantage = 1" is an indicator variable which is one when the email's subject line includes the corresponding comparative advantages. "March Chinese Virus = 1" is an indicator variable which is one when the email was sent between 03/18/2020-03/24/2020 when President Trump used the wording "Chinese Virus." "US Investor = 1" and "Female Investor = 1" are indicator variables for being a U.S. investor and being a female investor. Columns (1), (2), (3), and (5) use all the observations collected in the correspondence test. In Column (4), results are reported for the sub-sample where the startup team graduated from purely Ivy League colleges, Stanford and MIT. "Pure_Ivy" indicates cases like "Team from Columbia University" while "Mixed_Ivy" indicates cases like "Team from Columbia University and Juilliard Music School". For some startups in the music or medical industry, I combined an Ivy League college with a good university in that specific area for the treatment group. In Panel C, the dependent variable is the time spent on each pitch email measured in seconds. In Columns (1) and (2), I include unopened emails and replace their email staying time with 0 seconds. Considering the potential truncation issue, I also report the sub-sample of opened emails in Column (3). R^2 is the adjusted R^2 for all OLS regressions. Standard errors in parentheses are clustered at the investor level. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Experiment B Interaction Effects Based on Email Opening Rate

	Dependent Variable: 1(*Opened*)				
	(1)	(2)	(3)	(4)	(5)
	Full	"Mixed Ivy"	"Pure Ivy"	Full	After 03/24, "Pure Ivy"
Female Founder=1	0.006	0.002	0.009		
	(0.005)	(0.008)	(0.007)		
Asian Founder=1				0.009*	0.026***
				(0.005)	(0.008)
Ivy=1	0.003	-0.010	0.013*	0.010**	0.030***
	(0.005)	(0.008)	(0.007)	(0.005)	(0.008)
Ivy=1 × Female Founder=1	0.008	0.020*	-0.002		
	(0.007)	(0.011)	(0.010)		
Ivy=1× Asian Founder=1				-0.007	-0.032***
				(0.007)	(0.011)
US Investor=1	-0.016***	-0.009	-0.023***	-0.016***	-0.019**
	(0.006)	(0.008)	(0.008)	(0.006)	(0.009)
Female Investor=1	-0.019***	-0.023***	-0.017***	-0.019***	-0.011
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)
Constant	0.191***	0.191***	0.117***	0.190***	0.103***
	(0.019)	(0.020)	(0.013)	(0.019)	(0.013)
Pitch FE	Yes	Yes	Yes	Yes	Yes
Observations	30,909	14,331	16,578	30,909	13,006
R-squared	0.005	0.004	0.006	0.005	0.007

Notes. This table reports regression results of interaction effects between founders' educational backgrounds and founders' gender or race using investors' email opening rate as the outcome variable. The dependent variable is a dummy variable, which is one if an investor opens the pitch email, and zero otherwise. "Female Founder = 1" is an indicator variable that equals one if the first name of the email sender is a female name, and zero otherwise. Similarly, "Asian Founder = 1" is an indicator variable that equals one if the last name of the email sender is an East Asian name, and zero otherwise. "Ivy = 1" is an indicator variable for adding an Ivy League educational background in the email's subject line. "US Investor = 1" and "Female Investor = 1" are indicator variables for being a U.S. investor and being a female investor. To identify underlying dominant mechanisms, I include the interaction term of "Ivy = 1" and "Female Founder = 1" in Columns (1)-(3) and also the interaction term of "Ivy = 1" and "Asian Founder = 1" in Columns (4)-(5). Column (1) reports the regression results using all the observations in the correspondence test. In column (2), results are reported for the "Mixed Ivy" sub-sample, which indicates cases like "Team from Columbia University and Juilliard Music School." For some startups in the music or medical industry, I combine an Ivy League College with a good university in that specific area for the treatment group. In Column (3), results are reported for the "Pure Ivy" sub-sample, which indicates cases like "Team from Columbia University". The universities that founders have graduated from in the "Pure Ivy" cases are the Ivy League colleges, Stanford, and MIT. In Column (5), results are reported for the sub-sample where pitch emails are sent after 03/24 and emails belong to the "Pure Ivy" cases in order to increase the experiment's power. Note that President Trump stopped using "Chinese Virus" after 03/23/2020. R^2 is the adjusted R^2 for OLS regressions. Standard errors are in parentheses and are clustered at the investor level. *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance at 1%, 5%, and 10% levels, respectively.

Table 12: Experiment B Heterogeneous Effect of Investors' Response (Gender)

	Dependent Variable: 1(*Opened*)			Dependent Variable: *Email Staying time Unit:s*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Full	Female	Male	Full	Female	Male
Female Founder=1	0.011**	0.008	0.011**	0.147	0.301	0.076
	(0.004)	(0.007)	(0.004)	(0.996)	(1.884)	(0.997)
Female Founder=1 ×	-0.003			0.590		
Female Investor=1	(0.008)			(2.108)		
US Investor=1	–0.016***	-0.022*	-0.015**	1.229	-1.493	1.749
	(0.006)	(0.012)	(0.007)	(1.202)	(2.478)	(1.376)
Female Investor=1	-0.018***			-2.752*		
	(0.006)			(1.561)		
Constant	0.192***	0.153***	0.204***	24.124***	16.048***	26.432***
	(0.019)	(0.030)	(0.023)	(3.175)	(4.579)	(3.848)
Startup FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,909	7,277	23,632	3,720	767	2,953
R-squared	0.005	0.002	0.005	0.000	0.001	0.000

Notes. This table reports the heterogeneous effect of global investors' email opening behaviors in response to randomized pitch emails based on investors' gender in the correspondence test, which helps test the homophily mechanism. In Columns (1)-(3), the dependent variable is a dummy variable, which is one when an investor opens the pitch email, and zero otherwise. In Columns (4)-(6), the dependent variable is the time spent on each pitch email. In order to mitigate the truncation issue, I only include the opened emails in Columns (4)-(6). "Female Founder = 1" is an indicator variable that equals one if the first name of the email sender is a female name, and zero otherwise. Similarly, "Asian Founder = 1" is an indicator variable that equals one if the last name of the email sender is an East Asian name, and zero otherwise. "US Investor = 1" and "Female Investor = 1" are indicator variables for being a U.S. investor and being a female investor. R^2 is the adjusted R^2 for OLS regressions. Standard errors are in parentheses and are clustered at the investor level. *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance at 1%, 5%, and 10% levels, respectively.

Figures

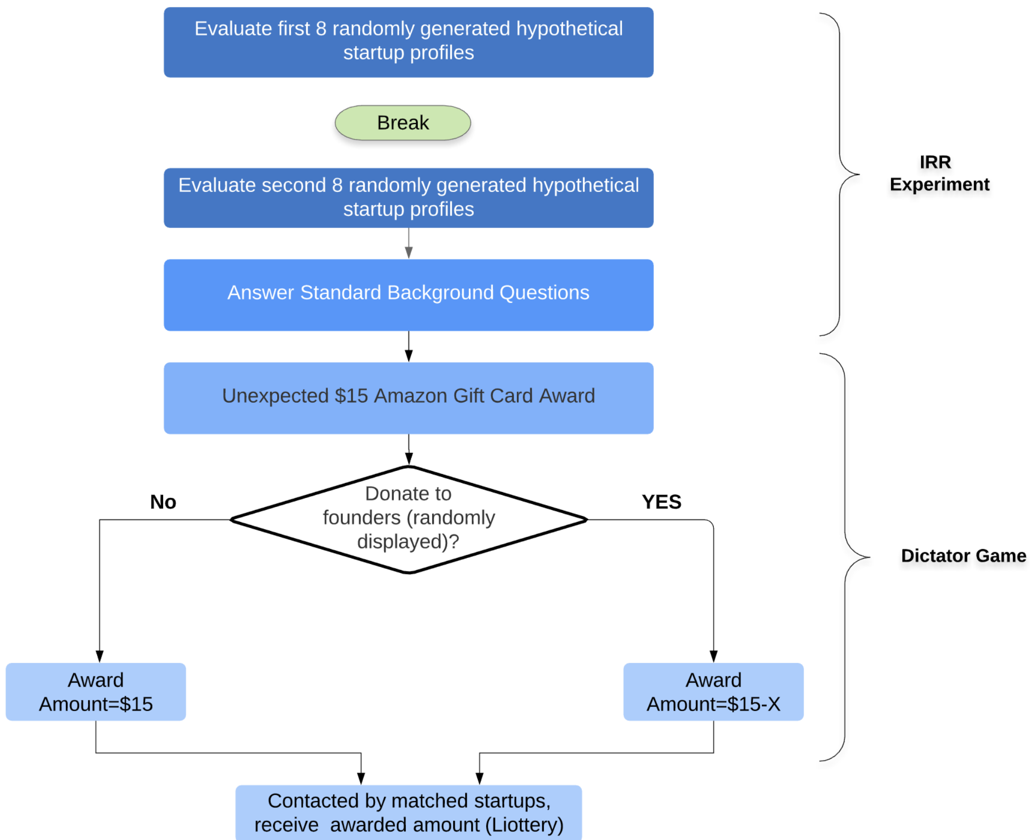


Figure 1: Experiment A Experimental Design

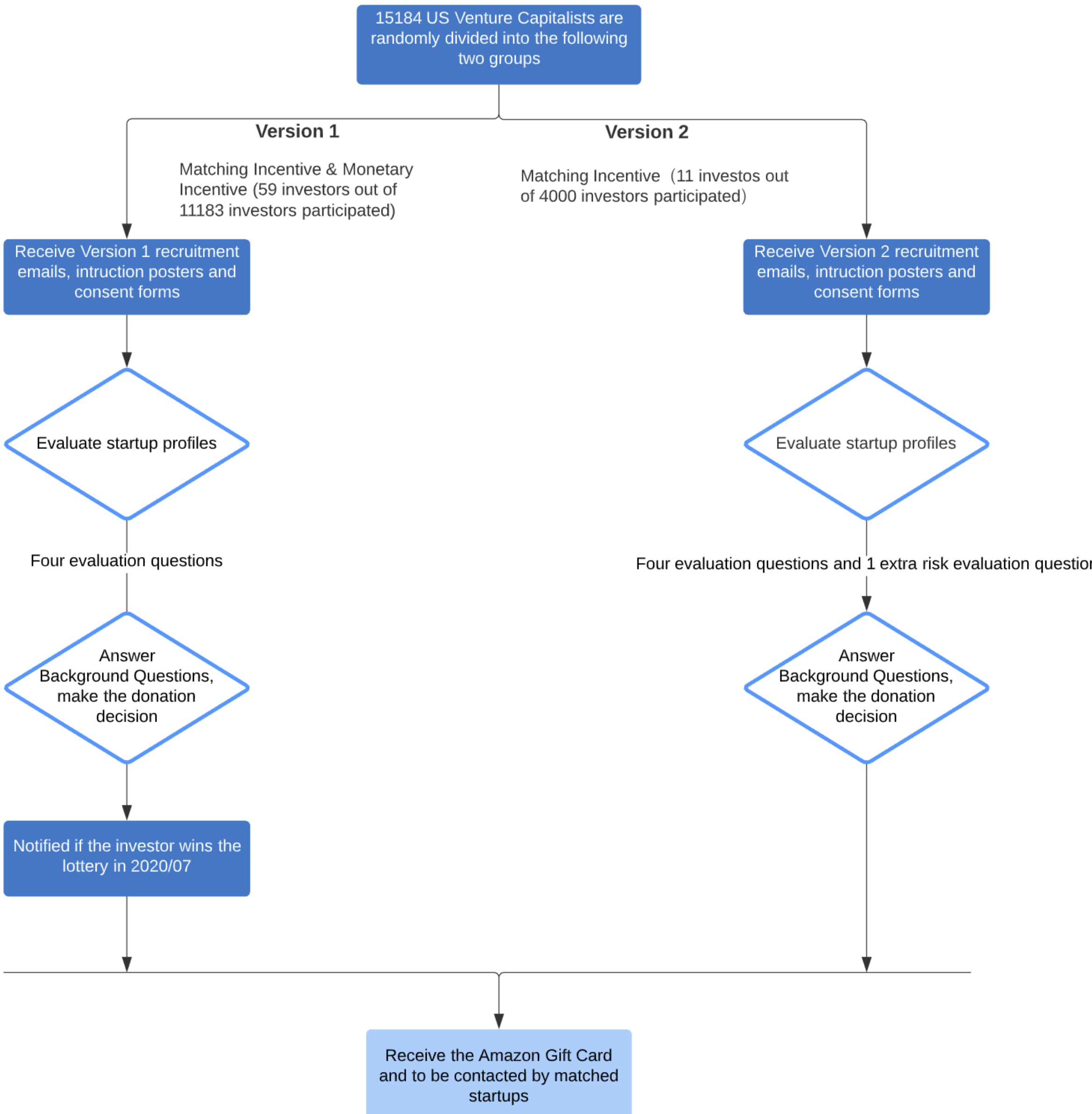


Figure 2: Experiment A Incentive Structure

Email Subject Line: Invest in StudioFinder {University1} {Advantage1}

Hello {Name},

My name is {FounderName}. I am the co-founder of StudioFinder, a music studio search app in New York City. Our team found your information in the VCPro Database, and we feel you might be interested in our startup.

StudioFinder is simply a music studio Airbnb that matches a studio holder and an artist. Our platform helps individuals who have a studio set in their house to make profits while they are not using it. StudioFinder provides new artists with affordable studio settings. The music studio rental corporations in New York cannot compete with our commission, which is the lowest (1.5 %), because we just match individuals. We have been collaborating with a few studios for a year and have had positive outcomes from new artists and video makers.

{Advantage2} {University2}

We are getting ready to raise funding to accelerate software adoption and bring StudioFinder to more users. If you are interested, we would love to share our pitch deck with you. Any feedback is also highly appreciated.

Thank you for your time. We look forward to your reply!

Sincerely,

{FounderName}

StudioFinder

Figure 3: Experiment B Example of a Pitch Email

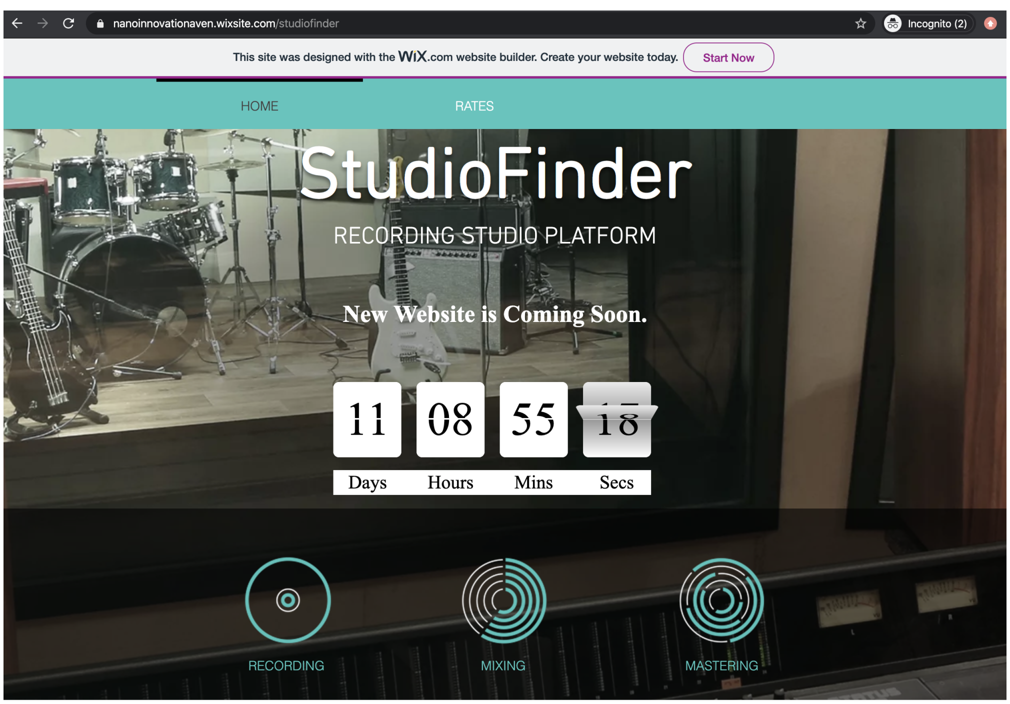


Figure 4: Experiment B Example of a Startup Website

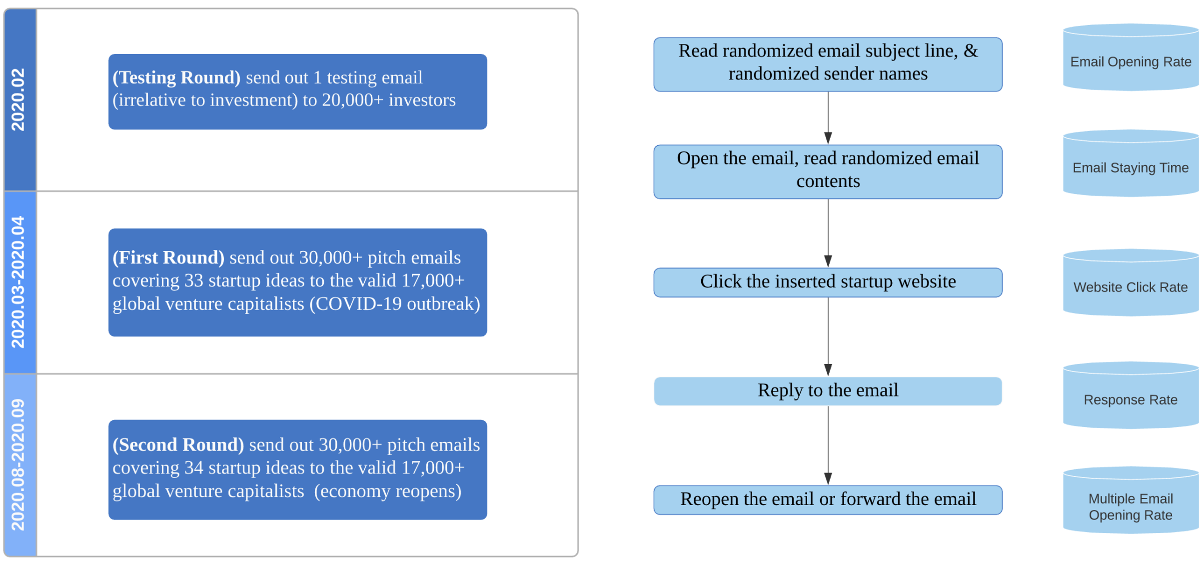


Figure 5: Experiment B Correspondence Test Experimental Design

Notes: This figure describes the experimental timeline, experimental design, and the tracked email behaviors of investors.

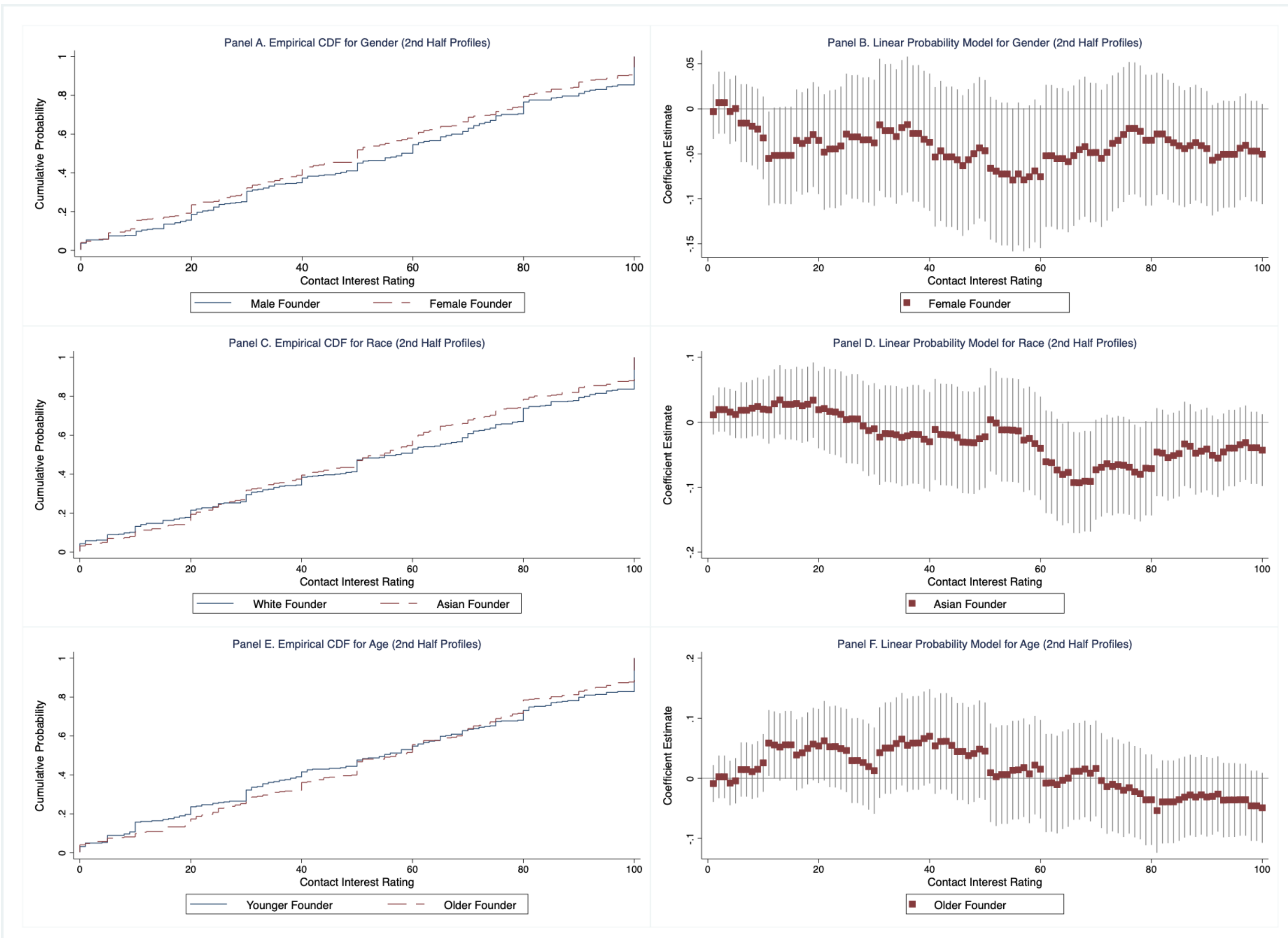


Figure 6: Effect of Founder's Gender, Race, and Age across the Contact Interest Distribution (2nd Half of Profiles)

Notes: This figure demonstrates the effect of startup founder's gender, race, and age across the contact interest distribution using the profiles evaluated in the second half of Experiment A. Panel A provides the empirical CDF for founder's gender on investors' contact interest rating (i.e., $Pr(\text{Contact Interest} \leq x|\text{Female Founder})$ and $Pr(\text{Contact Interest} \leq x|\text{Male Founder})$). Panel B provides the OLS coefficient estimates (i.e., $Pr(\text{Contact Interest} \leq x|\text{Female Founder}) - Pr(\text{Contact Interest} \leq x|\text{Male Founder})$) and the corresponding 95% confidence level. Similarly, Panels C and E provide the empirical CDF for founder's race and age. Panels D and F provide the OLS coefficient estimates for founder's race and age.

Online Appendix

A Data Construction Process

A.1 Data Sources

In order to construct an individual-level global venture capitalist database, containing both demographic information and contact information, I use the following commercial datasets as well as manually collected data.[92]

A.1.1 Pitchbook & CBInsight

The Pitchbook database contains extremely comprehensive information about venture capital and angel investors' demographic information and contact methods around the globe, though especially in the U.S. I purchased their individual-level data from 2017-2020 and selected the following types of investors from Pitchbook: Angel Group, Angel Individual Investor, Corporate Venture Capital, Family Office, and Venture Capital. CBInsight is used to complement the Pitchbook data.

A.1.2 ExactData

I also purchased a database of VC practitioners in the U.S. from the professional data company "ExactData. Inc", which collects the information from online websites and various VC industry events or gatherings. My research team verified and cleaned the database during the summer of 2018 and the spring of 2019, deleting those who have left the industry and correcting other invalid information. Moreover, we manually went through each firm contained in the database and added the contact information of new VC practitioners who were not contained in the original database through the following channels: personal websites, firm websites, LinkedIn, Zoominfo and Rocketreach.

A.1.3 SDC New Issue Database & Rocketreach

Rocketreach is one of the largest platforms and data sources providing contact information for company employees.[93] Given the company name list, it is feasible to extract the employees' contact information. Therefore, I implemented the following steps to further add investors' contact information:

Step 1: add new companies
I added many new venture capital funds to our previous database by checking the 2018 National Venture Capital Association (NVCA) member list and Thompson Reuters SDC Platinum VentureXpert Database.

Step 2: collect investors' information
Based on the fund list, I searched for all the employees working in the corresponding funds and companies using Rocketreach's API. I only kept the investment related positions, like VC investor, analysis, associate, VP, MD, etc. Rocketreach provided me with both the contact information (e.g. email and telephone number) and also the demographic information (e.g. Facebook, Twitter, LinkedIn, Position, etc.). For investors not contained in Pitchbook and ExactData, individual-level investors' demographic data were extracted manually from personal websites, Facebook, firm websites, LinkedIn, Zoominfo, and other social platforms.

A.1.4 Zdatabase

Zdatabase is provided by Zero2IPO Research Center and is currently one of the most comprehensive, accurate and timely databases covering the VC and PE industry in China.[94] It contains rich information about active Chinese

[92] Many of these commercial databases are not free and require researchers to sign a data contract for academic purposes. All key variables used in the analysis, including gender, location and industry, are manually verified through multiple social platforms including LinkedIn, company websites, personal websites and online news if such information is not available on Pitchbook.

[93] Using Rocketreach to collect contact information of employees is a very efficient data collection method. Given a company name list, researchers can extend the company level data to individual-level data by using Rocketreach. Potentially this data collection method can be implemented in a broad range of research in the labor economics and corporate finance field.

[94] Zdatabase description: http://www.p5w.net/fund/smjj/201209/P020120905327816063973.pdf

investment institutions and their management team starting from 1992. All the data are collected through regular surveys and daily phone calls, and are verified through many other available channels. The database is updated daily to provide an accurate, timely and authoritative data source. Considering that the research was implemented in English, I only included investors from Hong Kong and excluded investors from the Mainland.

A.2 Key variables

A.2.1 Gender

Pitchbook and ExactData contain each investor's gender information. For other investors not contained in these datasets, my research team manually verified their gender by searching online social platforms and company websites. For investors whose gender information is ambiguous, I excluded them from the recruitment list.

A.2.2 Location

Pitchbook and ExactData contain each investor's location information. For other investors not contained in these datasets, my research team manually collected their location information on LinkedIn or company websites.

A.2.3 Industry

Pitchbook contains each investor and their fund's detailed industry preferences. For other investors not contained in Pitchbook, my research team manually collected their individual-level preferences from LinkedIn and other social platforms. If the individual-level industry preferences are not available, I use the fund's industry preference instead. If no preference information is found online or from CBInsight or Pitchbook, I assume the investor does not have any specific investment preference. Such an assumption may result in extra noise and lower the email response rate in the correspondence test.

A.2.4 ESG

Pitchbook contains each fund's investment philosophy and their types. In the heterogeneous analysis based on a fund's ESG criteria, I treat those not-for-profit VC funds as impact funds and for-profit VC funds as common funds. This classification method potentially underestimates the fraction of ESG-related VC funds. An alternative way is to classify VC funds through selecting ESG-representative key words in their company description as Barber et al. (2020) did. However, the key word selection is very subjective and highly depends on context. Based on this more aggressive method, ESG-related funds can account for roughly 7% of the total observations. However, the basic heterogeneous effect analysis based on these two classification methods is similar.

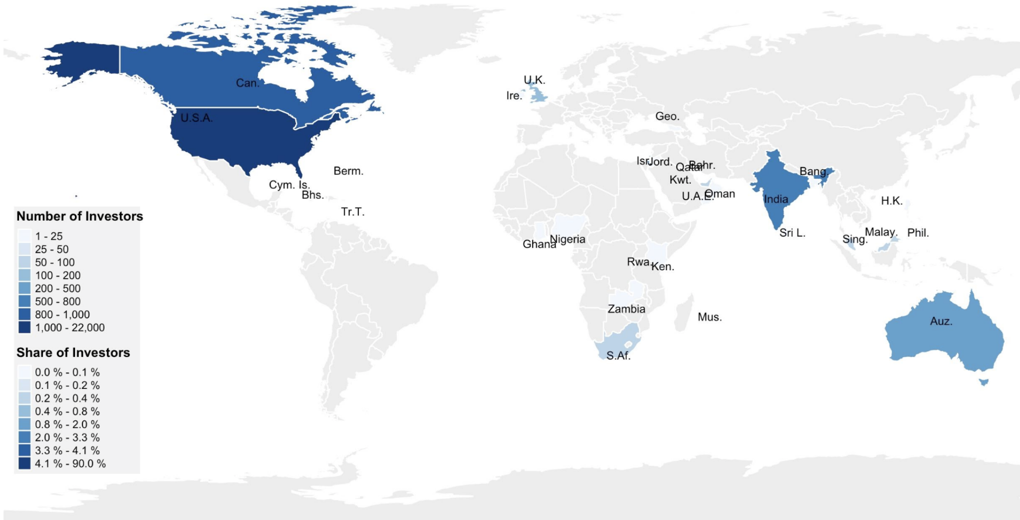


Figure A1: Geographical Distribution of Global Investors

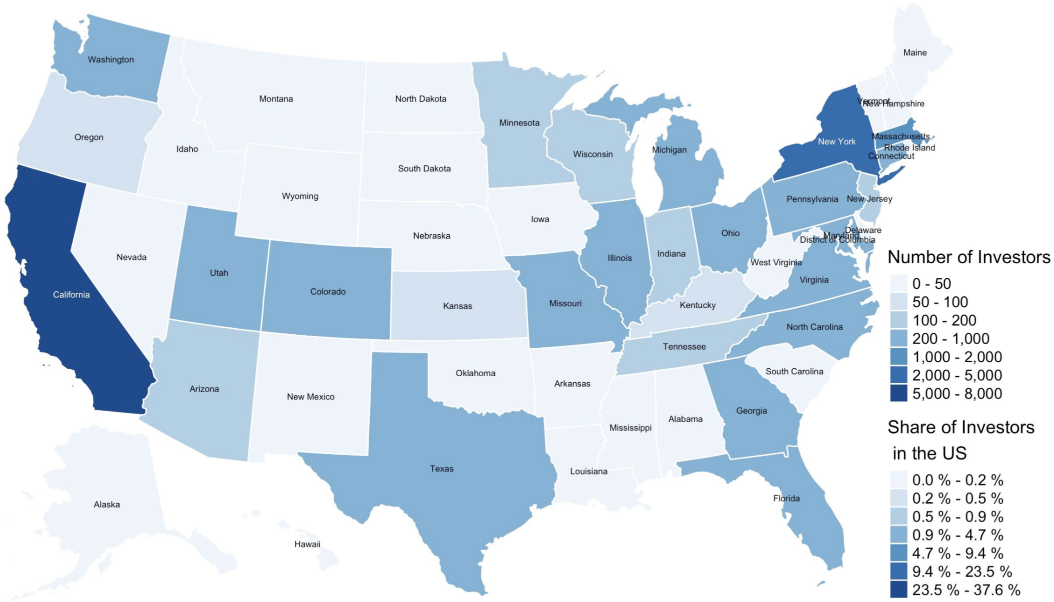


Figure A2: Geographical Distribution of U.S. Investors

B Lab-in-field Experiment

B.1 Startup Profile Construction Process

B.1.1 Startup Team Characteristics (Human Capital Assets)

Other related characteristics. —In addition to the gender, race, age, and educational background, I also randomize the following startup team characteristics, which are usually available on public platforms like LinkedIn, AngelList or CrunchBase. Such characteristics include the number of startup founders (1 or 2) and the founder's previous entrepreneurial experience. In order to accommodate investors from different industries, I use the wording "serial entrepreneur" to indicate the founding team's previous experiences.

B.1.2 Startup Project Characteristics (Non-human Capital Assets)

Comparative Advantages. —To indicate the quality of the startup project, I randomly generate a subset of common comparative advantages for the startups and use the number of these advantages to suggest the quality. However, considering that different comparative advantages are valued by investors from different industries,[95] I also asked investors which of the comparative advantages they would care about among the 10 comparative advantages used at the end of the tool and used the number of such cared about comparative advantages to confirm the results. The comparative advantage list is provided in Table B3.

Traction. —Traction is also an important indicator of the startup's financial situation and is measured by the previous monthly revenue and the annual revenue growth rate. Considering that we target early-stage investors, half of the startup profiles do not generate positive revenue yet and the other half have generated positive revenue. The range of the previous monthly revenue and the annual revenue growth rate comes from Pitchbook, which is biased towards more mature companies.[96]

Mission (ESG) —How ESG criteria affect investors' decisions is an important institutional question that has drawn more and more attention from both practitioners and researchers recently. In order to randomize the company's ESG criteria, I introduced a random variable called "Mission," which indicates whether such startups are purely profit driven (i.e. the control group, most commonly observed startups), profit driven with an IPO plan within 5 years (i.e. treatment 1 group), or also care about its environmental and social impact (i.e. treatment 1 group, social ventures). The description of the ESG-related mission is extracted from real social ventures.

Other related characteristics —Apart from the project characteristics mentioned above, I also added the following characteristics usually available on CrunchBase to enrich the startup profiles: startup founding date, company category (B2B or B2C),[97] number of employees, targeted market and location. Since the investors recruited in this experiment are U.S.-based investors, I only created two categories in terms of location, which includes the U.S. and outside the U.S., in order to test any potential home bias channels.

B.1.3 Previous Fund-raising Situation

Number of existing investors — Some investors may rely on previous investors' behaviors to make their decisions rather than relying on their own private information, especially when the previous investors are successful. Such herding behavior is documented in the IPO setting where subsequent investors ignore their private information and imitate earlier investors (Bikhchandani, Hirshleifer and Welch (1992)), and this is explained by informational cascades (Bikhchandani et al. (1992)). In order to test this behavior in the primary market, I also randomize the information of

[95] For example, investors in the tech industry may care more about registered intellectual properties in order to create entry barriers, while investors in the fashion industry may care more about celebrity endorsements rather than any tech-related advantages.

[96] The growth rate of some early stage startups can be 100% to 200% while most of the startups recorded in Pitchbook are have growth rates between 20% to 80%.

[97] Business to business or business to customers. These categories may affect investor's expectations since they are closely related to the startup's underlying business models. See the discussion on Tomasz Tungus' Twitter, who is an investor at Redpoint.

existing investors to indicate other investors' decisions similar to Bernstein et al. (2017). Existing investors' information is also available on multiple platforms like CrunchBase, Pitchbook or CB Insights. However, one limitation of such randomization is that I did not provide further background information of existing investors' financial backgrounds or reputation. Future researchers can provide more background information in order to better test this theoretical hypothesis.

Table B1: Experiment A Full Names Populating Profile Tool

Asian Female	**White Female**	**Asian Male**	**White Male**
Cynthia Huynh	Amber Morris	Evan Liu	Patrick Kelly
Jennifer Tang	Erica Carpenter	Alan Wu	Stephen Bennett
Amanda Cheung	Anna Hoffman	Bryan Liang	Steven Martin
Christina Chang	Amanada Gray	William Chung	Jeremy White
Linda Chung	Tiffany Roberts	Nicholas Wang	Jason Adams
Brittany Yi	Lisa Taylor	Charles Luu	Donald Schultz
Megan Ho	Karen Carroll	Zachary Ho	Jack Wright
Emily Xu	Danielle Collins	Marcus Yoon	Victor Becker
Jacqueline Lin	Megan Bennett	George Thao	Michael Hughes
Kayla Wang	Brenda Cox	Vincent Huynh	Keith Meyer
Cassandra Kwon	Kathleen Phillips	Luke Yang	Anthony Roberts
Julie Chan	Amber Sullivan	Justin Dinh	Justin Cooper
Monica Luong	Madeline Walsh	Matt Hoang	Benjamin Hill
Amber Hoang	Abigail Kelly	Jacob Xu	Mark Myers
Sara Truong	Alicia Cook	Donald Choi	Phillip Baker
Katrina Tsai	Amanda Jensen	Dennis Lin	Vincent Peterson
Abigail Zhao	Angela Larson	Victor Kwon	Dennis Reed
Vanessa Choi	Hayley Thompson	Jason Pham	Frank Phillips
Patricia Li	Christine Campbell	Eric Duong	Shane Taylor
Lisa Zhou	Caroline Parker	Stephen Hsu	William Welch
Caroline Lu	Kristy Baker	Kevin Jiang	Bryan Ward
Melissa Hwang	Tina Reed	Jeffrey Chen	Ian Russell
Mary Pham	Sara Burke	Erik Luong	Brian Wilson
Amy Hu	Victoria Snyder	Philip Zhao	Seth Schwartz
Jenna Nguyen	Molly Weaver	Jeremy Yu	Jared Walsh
Margaret Liang	Melissa Stone	Seth Truong	Zachary Parker
Danielle Liu	Melanie Wilson	Ian Zhou	John Carpenter
Megan Dinh	Rachael Ward	Matthew Chang	Jeffery Cook
Melanie Yang	Elizabeth Miller	Scott Lu	Nathan Nelson
Amanda Thao	Mary Hill	Sean Hwang	Matthew Rogers
Sarah Yu	Amy Moore	Patrick Hu	George Barker
Nichole Liu	Vanessa Smith	Mark Chan	Sean Beck
Christine Cho	Teresa Anderson	Jack Zhu	David Hall
Victoria Xiong	Catherine Schultz	Timothy Cheng	Andrew Miller
Teresa Wong	Heather Martin	Benjamin Nguyen	Peter Keller
Kara Yoon	Kathryn Myers	Steven Tang	Luke Jensen

Continued

Asian Female	**White Female**	**Asian Male**	**White Male**
Kathleen Cheng	Katie Meyer	Travis Wong	Kevin Hansen
Angela Wu	Valerie Price	David Zheng	Dustin Sullivan
Catherine Zheng	Melinda Evans	Paul Ngo	Philip Morris
Hayley Huang	Sandra Wright	Anthony Yi	Evan Moore
Karen Ngo	Christina Russell	Shane Huang	Paul Burke
Elizabeth Duong	Kayla Allen	Robert Zhang	Matt Price
Laura Luu	Jacqueline Schmidt	Kenneth Tsai	Marcus Collins
Rebecca Hsu	Jennifer Welch	Richard Xiong	Richard Thompson
Melinda Zhang	Michelle Nelson	Brian Cho	Thomas Snyder
Katherine Le	Sarah Fisher	Joel Le	Christopher Larson
Tara Jiang	Brittany Rogers	Michael Li	Travis Gray
Alicia Zhu	Grace Keller	Trevor Cheung	Charles Hoffman
Molly Huynh	Julie Beck	Adam Liu	Joel Stone
Samantha Tang	Monica Cooper	Peter Wu	Joseph Allen

Notes. This table provides the name lists of hypothetical startup founders used in the survey tool. 50 names were selected to be highly indicative of each combination of race and gender. Considering the White and Asian startup founders account for most of the highly innovative startups, we only have four combinations listed above: Asian Female, White Female, Asian Male, White Male. A name drawn from these lists is displayed at the beginning part of the startup profiles and in the questions used to evaluate the resumes. First and last names are linked every time they appeared, and the combinations of first and last names are randomly generated. Considering that Asian and White Americans have very similar naming patterns as documented by Fryer Jr and Levitt (2004), I choose their first names from the same name pool. After I generated a list of potential full name candidates, I further removed those names owned by famous startup founders or CEOs. That's why there are slight differences between first names for Asian founders and first names for white founders. Names were selected uniformly and without replacement within the chosen column of the table. I use the variation induced by these names for the analysis variables Female, Asian; Female, White; Male, Asian; Male, White. I did not list the gender information explicitly, as the Crunchbase platform does (For example, by adding one more bullet point: Gender: Male), due to the experiment observer effect.

Table B2: Experiment A Educational Background (School List)

School Category	Universities	Percentage
(Top School)	Brown University	50%
Example	Columbia University	
	Cornell University	
	Dartmouth College	
	Harvard University	
	Princeton University	
	University of Pennsylvania	
	Yale University	
	California Institute of Technology	
	MIT	
	Northwestern University	
	Stanford University	
	University of Chicago	
(Common School)	Thomas Jefferson University(153)	50%
Example	University of Arkansas(153)	
	Hofstra University(162)	
	University of Mississippi (162)	
	Virginia Commonwealth University (162)	
	Adelphi University (166)	
	University of Maryland-Baltimore County(166)	
	University of Rhode Island(166)	
	St.John's University (179)	
	University of Detroit Mercy (179)	
	University of Idaho (179)	
	Biola University (185)	
	Chatham University (185)	
	Bellarmine University (197)	
	Bethel University (197)	
	Loyola University New Orleans (197)	
	Robert Morris University (202)	
	Regis University (202)	
	Widener University(202)	
	Laurentian University (Canada)	
	Auburn University (104)	
	Rochester Institute of Technology (104)	
	University of Tulsa (121)	
	DePaul University (125)	

Notes. This table provides the school list used to generate the educational background of each hypothetical startup founder. The percentage of top school and common school is 50% vs. 50% to increase the power. Also, for highly innovative startups, their founders are more likely to have graduated from prestigious universities. Top schools refer to the Ivy League schools (Brown University, Columbia University, Cornell University, Dartmouth College, Harvard University, Princeton University, University of Pennsylvania, and Yale University) as well as other top U.S. schools (Amherst College, California Institute of Technology, Duke University, MIT, Northwestern University, Stanford University, University of California, Berkeley, University of Chicago, and Williams College). Since the incubators that we collaborate with have more connections with Columbia University and Stanford University, we give more weight to these universities. Common Schools are those ranked lower than the 150th based on the U.S. News 2020 ranking results. I also add a Canadian common school since one of the incubators is from Canada.

Table B3: Experiment A Company Comparative Advantage

Advantage Category	Description
(Product)	trade secrets/patents registered
	celebrity endorsement
	exclusive partnerships
	accumulated many pilot consumers
	adoption of the latest technology
	pricing advantage
	great product design
	1st mover
(Cost)	lower cost
	economies of scale
Total	100%

Notes. I use the number of the corresponding comparative advantages as a measure of the quality of the startup project. For each startup profile, the subset of comparative advantages is randomly drawn from the 10 advantages listed above.

Table B4: Experiment A Design, Incentives Design for Different Evaluation Questions

Evaluation Questions	Matching Incentive (Version 2)	Monetary Incentive	Matching and Monetary Incentive (Version 1)
Q1 (quality evaluation)	Yes	Yes	Yes
Q2 (collaboration likelihood)	Yes	No	Yes
Q3 (contact interest)	Yes	No	Yes
Q4 (investment interest)	Yes	No	Yes
Q5 (risk evaluation)	Yes	N/A	Yes

Notes. This tables describes how different types of incentives affect each evaluation question. Column 1 shows that the matching incentive, which identifies the matched startups using the matching algorithm, works for all five of the evaluation questions. I sent Version 2 recruitment emails, instruction posters, and consent forms to investors who only receive this matching incentive. Column 2 shows that the monetary incentive, which provides a lottery opportunity, only incentivizes Q1 (the evaluation of the startup quality evaluations) because the financial returns for the lottery winners only depends on the belief of the startup's financial return. Column 3 shows that combining the matching and the monetary incentive together can also incentivize all five questions. I sent Version 1 recruitment emails, instruction posters, and consent forms to investors who received both incentives.

Table B5: Experiment A Evaluation Results (Team vs. Project)

Dependent Variable	Q1 Quality (1)	Q2 Collaboration (2)	Q3 Contact (3)	Q3 Contact (4)	Q4 Investment (5)	Q4 Investment (6)	Q5 Risk (7)
Serial Founder	5.23***	-0.81	5.64***	1.26	0.76***	0.13	-0.65
	(1.08)	(0.88)	(1.28)	(0.91)	(0.19)	(0.15)	(3.05)
Ivy	5.36***	-1.06	7.44***	3.01***	0.87***	0.20	-6.44**
	(1.10)	(0.87)	(1.31)	(0.93)	(0.20)	(0.15)	(3.26)
Number of Founders	1.56	-1.21	1.17	-0.11	0.21	0.04	-5.32*
	(1.07)	(0.88)	(1.29)	(0.91)	(0.20)	(0.15)	(3.06)
US Founder	0.95	0.02	4.23***	3.69***	0.08	0.03	-0.91
	(1.18)	(0.91)	(1.39)	(1.00)	(0.21)	(0.16)	(3.48)
# Comparative Adv	3.10***	-0.22	2.76***	0.34	0.55***	0.15**	0.91
	(0.54)	(0.43)	(0.64)	(0.43)	(0.10)	(0.07)	(1.48)
Has Positive Traction	12.70***	1.75**	13.35***	1.91*	1.81***	0.28*	-9.51***
	(1.07)	(0.86)	(1.28)	(0.99)	(0.20)	(0.16)	(3.15)
Number of Employees [0-10]	0.67	2.37**	-1.73	-2.57**	-0.19	-0.29	-1.18
	(1.43)	(1.16)	(1.69)	(1.18)	(0.26)	(0.20)	(3.94)
Number of Employees [10-20]	-1.08	0.94	-3.26	-2.08	-0.46	-0.33	
	(1.64)	(1.35)	(1.99)	(1.39)	(0.30)	(0.23)	
Number of Employees [20-50]	-0.47	-0.02	-1.21	-0.72	-0.16	-0.12	-1.28
	(1.45)	(1.17)	(1.71)	(1.17)	(0.27)	(0.19)	(3.59)
Company Age	-4.59*	-5.99***	-7.39**	-2.19	-1.26**	-0.54	-3.41
	(2.72)	(2.19)	(3.19)	(2.26)	(0.49)	(0.37)	(7.74)
Company Age^2	0.75	1.12**	1.27**	0.42	0.23**	0.10	0.77
	(0.54)	(0.44)	(0.64)	(0.45)	(0.10)	(0.07)	(1.52)
Is B2B	3.90***	3.73***	6.10***	1.47	0.81***	0.32**	-4.91
	(1.07)	(0.86)	(1.28)	(0.89)	(0.20)	(0.15)	(3.01)
Domestic Market	-0.10	-0.60	0.09	0.57	0.08	0.13	-3.32
	(1.08)	(0.86)	(1.28)	(0.90)	(0.20)	(0.14)	(3.19)
Q1				0.88***		0.12***	
				(0.03)		(0.01)	
Q2				0.18***		0.01	
				(0.03)		(0.01)	
Constant	49.75***	78.20***	66.20***	-4.19	5.62***	-0.33	67.01***
	(6.56)	(6.02)	(4.93)	(7.50)	(1.43)	(0.63)	(11.66)
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	1,216	1,184	1,216	1,184	1,176	1,154	176
R-squared	0.44	0.55	0.56	0.80	0.44	0.70	0.34

Notes. This table shows that investors understand the incentives and care about multiple important startup team and project characteristics. In columns (1)-(7), the dependent variable is the evaluation results of Q1 (quality evaluation), Q2 (collaboration interest), Q3 (contact interest), Q3 (contact interest), Q4 (contact interest), Q4 (investment interest), and Q5 (risk evaluation). "Serial Founder," "Ivy," "US Founder," "Has Positive Traction," "Is B2B," and "Domestic Market" are indicative variables that equal to one if the founder is a serial entrepreneur, graduates from an Ivy League college, or lives in the U.S., and the project has positive traction, is a business-to-business startup, or focuses on the domestic market. These variables are equal to 0 if the startup does not have such characteristics. "Number of founders" is either 1 or 2; "Number of Comparative Advantages" and "Company Age" can be {1,2,3,4}; "Company Age^2" is the square of the company age. "Q1" is the evaluation results of startup quality. "Q2" is the evaluation results of the collaboration likelihood. All the regression results add investor fixed effect and use the robust standard errors reported in parentheses. I use the Bonferroni method to implement multiple hypothesis testing. *** p<0.01, ** p<0.05, * p<0.1

Table B6: Experiment A Incentive Structure Comparison

	(1) Q1	(2) Q2	(3) Q3	(4) Q4
Panel A: Gender				
Female Founder	-0.60	0.57	-0.34	0.02
	(1.29)	(0.99)	(1.53)	(0.23)
Female Founder × Matching	0.30	-0.77	-4.18	0.13
	(3.39)	(2.19)	(4.02)	(0.59)
Matching	-13.80	48.13***	15.28***	-0.87
	(9.58)	(3.93)	(2.61)	(1.76)
Investor FEs	Yes	Yes	Yes	Yes
Observations	1,216	1,184	1,216	1,176
R-squared	0.31	0.53	0.47	0.34
Panel B: Race				
Asian Founder	-0.28	-0.61	-0.75	-0.18
	(1.29)	(0.99)	(1.51)	(0.23)
Asian Founder × Matching	2.26	0.03	2.81	0.93
	(3.40)	(2.26)	(4.11)	(0.58)
Matching	-14.78	47.73***	11.78***	-1.26
	(9.84)	(3.97)	(2.57)	(1.75)
Investor FEs	Yes	Yes	Yes	Yes
Observations	1,216	1,184	1,216	1,176
R-squared	0.31	0.53	0.47	0.34
Panel C: Age				
Age	-0.46	-0.35	-0.43	-0.06
	(0.49)	(0.38)	(0.57)	(0.09)
Age^2	0.00	0.00	0.00	0.00
	(0.01)	(0.00)	(0.01)	(0.00)
Age × Matching	2.64**	0.75	0.63	0.33
	(1.34)	(0.85)	(1.58)	(0.23)
Age^2 × Matching	-0.03*	-0.01	-0.01	-0.00
	(0.02)	(0.01)	(0.02)	(0.00)
Matching	-54.95*	15.71	-37.53	-5.48
	(28.57)	(18.56)	(32.35)	(4.94)
Investor FEs	Yes	Yes	Yes	Yes
Observations	1,216	1,184	1,216	1,176
R-squared	0.31	0.53	0.47	0.34

Notes. This table compares the evaluation results of investors who are recruited by the following two incentive structures: "matching incentive + monetary incentive" and the "matching incentive" only. "Matching" is an indicator that equals to 1 when only the matching incentive is provided in the recruitment process, and zero otherwise. Panel A shows the comparison of evaluation results related to a founder's gender. Panel B shows the comparison of evaluation results related to a founder's race. Panel C shows the comparison of evaluation results related to a founder's age. Column (1) shows the Q1 (quality evaluation) regression. Column (2) shows the Q2 (collaboration likelihood) regression. Column (3) shows the Q3 (contact interest) regression. Column (4) shows the Q4 (investment interest) regression. All regression specifications add fixed effects for each investor. *** p<0.01, ** p<0.05, * p<0.1.

Table B7: Characteristics of the 1^{st} Half Profiles & Investors' 2^{nd} Half Evaluations

	(1) Q1	(2) Q2	(3) Q3	(4) Q4
Fraction of Female Founders	17.77	21.80	33.60	0.94
In the first Half	(17.03)	(17.32)	(22.39)	(2.76)
Fraction of Asian Founders	0.29	2.47	-30.84*	-0.34
In the first Half	(10.95)	(13.94)	(16.44)	(1.95)
Fraction of Older Founders	-1.02	2.02*	-0.77	-0.10
In the first Half	(1.40)	(1.20)	(2.06)	(0.26)
Observations	68	68	68	67

Notes. This table tests whether investors' second half evaluation ratings of the minority founders decrease when they evaluate more minority founders in the first half profiles. The dependent variable "Q1" "Q2" "Q3" "Q4" represents the average rating of "Q1" "Q2" "Q3" "Q4" in the second half of Experiment A. "Fraction of Female Founders In the first Half", "Fraction of Asian Founders In the first Half" and "Fraction of Older Founders In the first Half" stand for the fraction of female founders, Asian founders and older founders in the first half profiles respectively. These cross sectional regressions use robust standard errors. One investor participated in the experiment twice, so we delete his responses. However, results are still robust after including his responses.

Table B8: Compare the Performance of Women-led and Men-led Startups

	Raised New Funding		Out of Business		IPO/Acquisition	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Global Startups						
All Female Founders	-0.049***	-0.009	0.009**	0.003	-0.005***	-0.002
	(0.008)	(0.008)	(0.004)	(0.004)	(0.001)	(0.001)
Mixed Gender Founders	0.003	0.017***	-0.005*	-0.006**	-0.002	-0.001
	(0.005)	(0.005)	(0.003)	(0.003)	(0.001)	(0.001)
Observations	44,215	44,215	44,215	44,215	44,215	44,215
R-squared	0.06	0.17	0.06	0.07	0.05	0.06
Control	No	Yes	No	Yes	No	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Stage FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel B. U.S. Startups						
All Female Founders	-0.031***	0.010	0.012*	0.004	-0.008***	-0.003
	(0.012)	(0.011)	(0.007)	(0.007)	(0.002)	(0.003)
Mixed Gender Founders	0.002	0.024***	-0.005	-0.008**	-0.003	-0.001
	(0.008)	(0.007)	(0.004)	(0.004)	(0.003)	(0.003)
Observations	17,852	17,852	17,852	17,852	17,852	17,852
R-squared	0.05	0.19	0.05	0.07	0.04	0.06
Control	No	Yes	No	Yes	No	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Stage FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel C. IT Industry						
All Female Founders	-0.056***	-0.015	0.031***	0.024***	-0.001	0.001
	(0.015)	(0.015)	(0.009)	(0.009)	(0.004)	(0.004)
Mixed Gender Founders	-0.006	0.015**	-0.006	-0.008	-0.002	-0.002
	(0.008)	(0.008)	(0.005)	(0.005)	(0.002)	(0.002)
Observations	18,539	18,539	18,539	18,539	18,539	18,539
R-squared	0.06	0.19	0.08	0.09	0.05	0.06
Control	No	Yes	No	Yes	No	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Stage FE	Yes	Yes	Yes	Yes	Yes	Yes

	Raised New Funding		Out of Business		IPO/Acquisition	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel D. Early Stage						
All Female Founders	-0.045***	-0.007	0.011**	0.003	-0.003**	-0.002
	(0.009)	(0.008)	(0.005)	(0.005)	(0.001)	(0.001)
Mixed Gender Founders	0.010*	0.023***	-0.006*	-0.006*	0.000	0.000
	(0.006)	(0.005)	(0.003)	(0.003)	(0.001)	(0.002)
Observations	31,962	31,962	31,962	31,962	31,962	31,962
R-squared	0.06	0.19	0.06	0.08	0.05	0.05
Control	No	Yes	No	Yes	No	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table tests whether women-led ventures underperform men-led ventures during the 1 year period after the experiment (i.e., 2020/07/31-2021/07/31). The sample contains all the startups which has received funding between 2017/01/01 and 2020/07/31 and whose founders' gender information is observable in the Pitchbook data. Panel A examines the performance of global startups. Panel B focuses on the performance of only U.S. startups, defined as startups whose headquarters are located in the US. Panel C zooms into the IT-related startups. Panel D discusses startups whose latest financing round is still in the early stage or seed stage. In columns (1) and (2) of each panel, the dependent variable is an indicator equal to one if the startup either successfully raised new funding from the venture capital industry or the deal is in progress during 2020/07/31-2021/07/31. In columns (3) and (4), the dependent variable is an indicator equal to one if the startup's business status is "out of business" in 2021/10. Ideally I should use the business status in 2021/07/31, however, this information is not available to me. "out of business" is defined as either "File Bankruptcy" or "Out of Business" in Pitchbook. Results are still robust when including cases where the startup's website does not function anymore, such as reporting a 404 error. In columns (5) and (6), the dependent variable is an indicator equal to one if the startup filed an IPO or was acquired between 2020/07/31 and 2021/07/31. Columns (2), (4), and (6) include the following control variables which describe the last updated startup characteristics before 2020/07/31: number of deals, founding years, log (1+raised amount of the latest deal). Robust standard errors clustered at the headquarter location level are reported in parentheses. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B9: Rule Out Attention Discrimination In the Profile Selection Process

Dependent Variable	Response Time (1)	Response Time (2)	Response Time (3)
Second Half of Study	-25.65***	-28.11***	-26.53***
	(3.37)	(3.06)	(3.31)
Female Founder	0.20		
	(3.83)		
Female Founder × Second Half of Study	-3.10		
	(4.73)		
Asian Founder		-0.42	
		(3.86)	
Asian Founder × Second Half of Study		1.93	
		(4.83)	
Older Founder			3.11
			(3.77)
Older Founder × Second Half of Study			-1.37
			(4.66)
Investor FE	Yes	Yes	Yes
Observations	1216	1216	1216
R-squared	0.34	0.34	0.34

Notes. This tables tests whether investors' response time decreases for minorities in the second half of the study (i.e., "attention discrimination"). The dependent variable is investors' response time, which is defined as the number of seconds before each page submission, winsorized at the 95th percentile (59.23 seconds on average). "Female Founder" is an indicator equal to one if the startup founder has a female first name, and zero otherwise. "Asian Founder" is an indicator equal to one if the startup founder has an Asian last name, and zero otherwise. "Older Founder" is an indicator equal to one if the startup founder graduated from college in 2005 or before, and zero otherwise. "Second Half of Study" is an indicator variable for startup profiles shown among the last eight resumes viewed by a subject. Standard errors in parentheses are robust standard errors. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

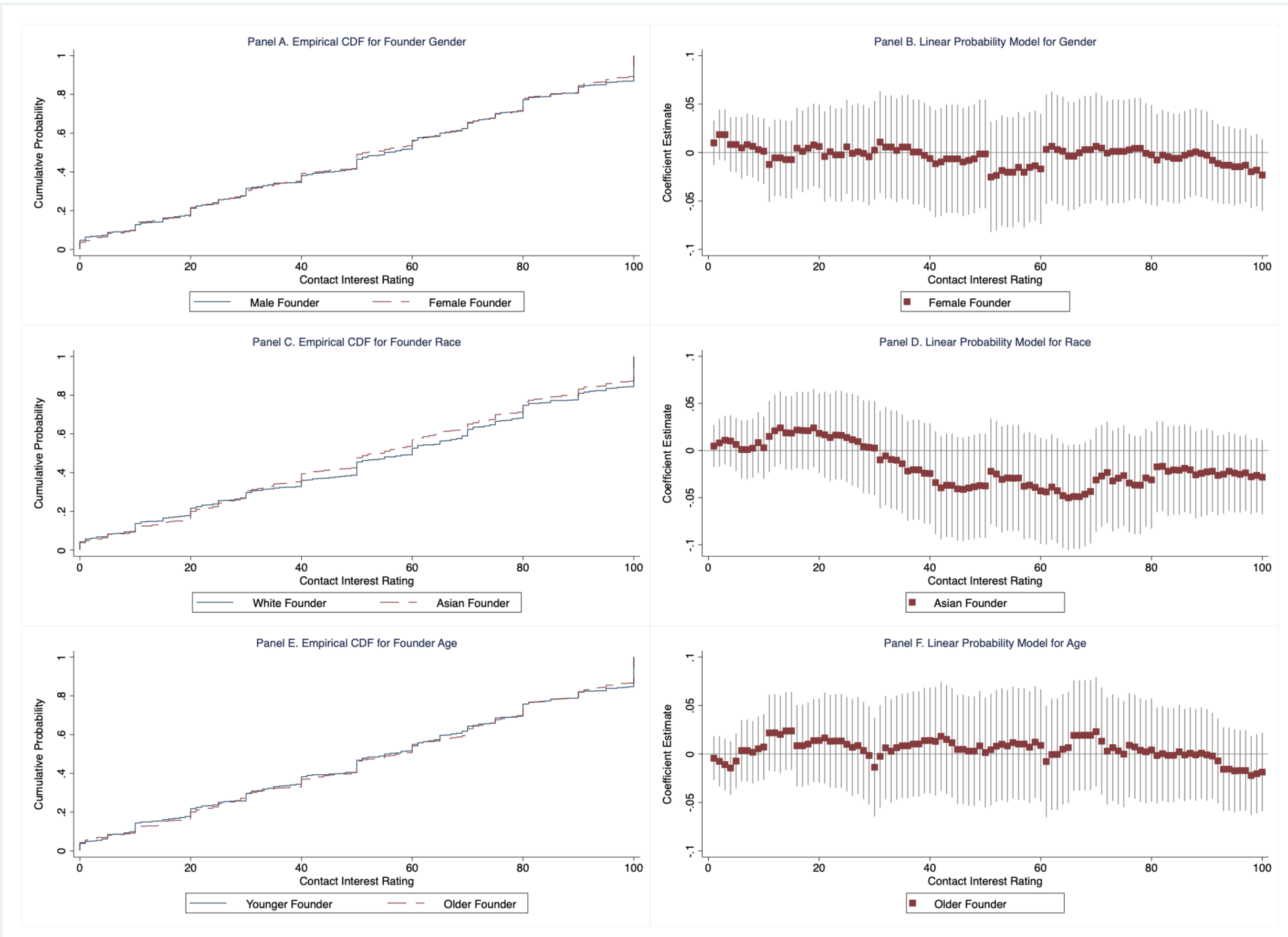


Figure B3: Effect of Founder's Gender, Race, and Age across the Contact Interest Distribution (Total Profiles)

Notes: This figure demonstrates the effect of a startup founder's gender, race, and age across the contact interest distribution using the total profiles evaluated in Experiment A. Panel A provides the empirical CDF for a founder's gender on investors' contact interest rating (i.e. $Pr(\text{Contact Interest} \leq x|\text{Female Founder})$ and $Pr(\text{Contact Interest} \leq x|\text{Male Founder})$). Panel B provides the OLS coefficient estimates (i.e. $Pr(\text{Contact Interest} \leq x|\text{Female Founder}) - Pr(\text{Contact Interest} \leq x|\text{Male Founder})$) and the corresponding 95% confidence level. Similarly, Panels C and E provide the empirical CDF for a founder's race and age. Panels D and F provide the OLS coefficient estimates for a founder's race and age.

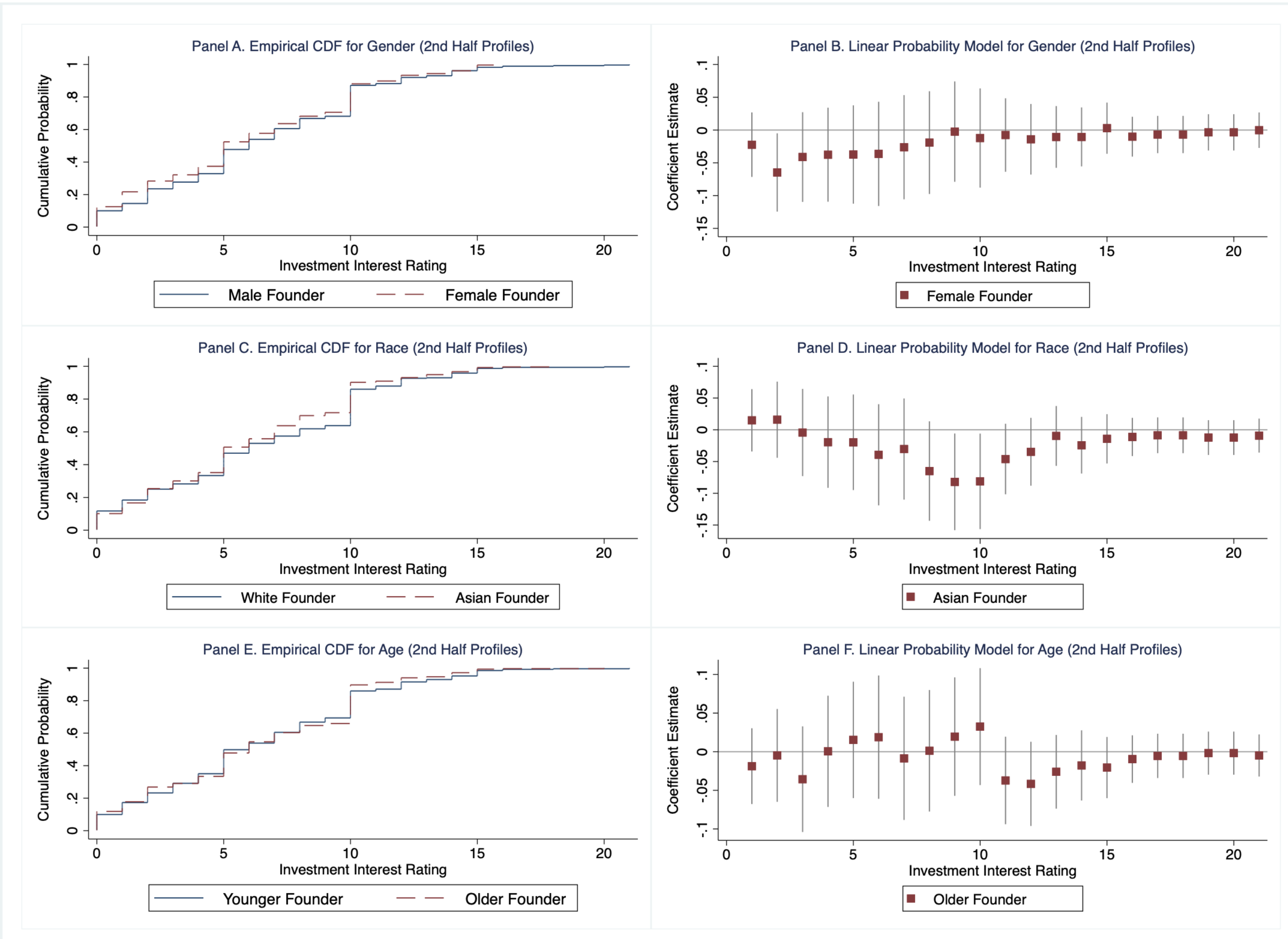


Figure B4: Effect of Founder's Gender, Race, and Age across the Investment Interest Distribution (2nd Half of Profiles)

Notes: This figure demonstrates the effect of startup founder's gender, race, and age across the investment interest distribution using the profiles evaluated in the second half of Experiment A. Panel A provides the empirical CDF for founder's gender on investors' investment interest rating (i.e. $Pr(\text{Investment Interest} \leq x|\text{Female Founder})$ and $Pr(\text{Investment Interest} \leq x|\text{Male Founder})$). Panel B provides the OLS coefficient estimates (i.e. $Pr(\text{Investment Interest} \leq x|\text{Female Founder}) - Pr(\text{Investment Interest} \leq x|\text{Male Founder})$) and the corresponding 95% confidence level. Similarly, Panels C and E provide the empirical CDF for founder's race and age. Panels D and F provide the OLS coefficient estimates for founder's race and age. There are two key differences between investment interest ("Q4") and contact interest ("Q3"). First, investment interest ("Q4") is noisier than contact interest ("Q3") because this experiment does not provide any soft information about the startup founder, which is also crucial to investment decisions. Second, the investment interest is mainly affected by beliefs rather than tastes as documented by Zhang (2020). Therefore, when using investment interest to illustrate the distributional effect (see Appendix B Figure B4), the discrimination reversion pattern is slightly noisier for Asian founders and older founders, and disappears for female founders. The reason is because any taste-driven preference towards women (proved in Experiment B) mainly plays a role in the contact stage rather than the investment stage. Therefore, these results are expected according to Zhang (2020).

Startup Team Evaluation Section

Instructions:

All 16 startup teams are hypothetical and randomly generated. However, we will help you find real high-quality startup teams, which have connections with our collaborative incubators, based on your choices and ratings in this survey. The matched startup teams will contact you after 1 month.

We will use all evaluation answers to recommend highly matched startup teams from our collaborative incubators. All data will be kept strictly confidential and analyzed at the aggregate level after removing identifiable information.

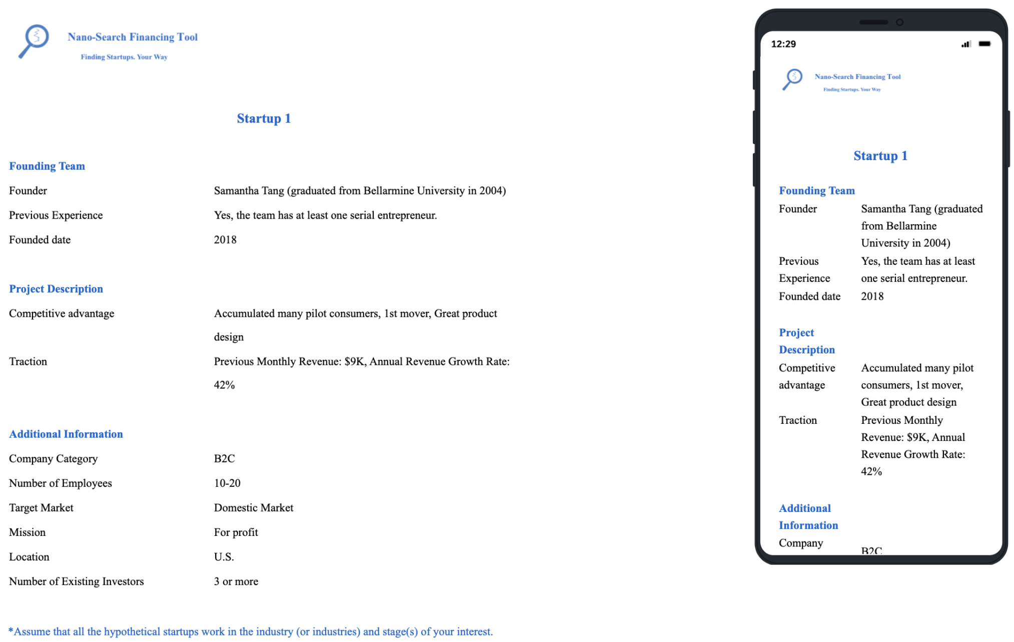
Note:

1. Assume that all the hypothetical startups work in the industry (or industries) and stage(s) of your interest and that all startup teams have adequate knowledge of the industry.

2. The more carefully and truthfully you evaluate each startup profile, the more benefits you can get.

Back | Next

Figure B5: Experiment A Instruction Page (Version 2)



Nano-Search Financing Tool
Finding Startups. Your Way

Startup 1

Founding Team

Founder	Samantha Tang (graduated from Bellarmine University in 2004)
Previous Experience	Yes, the team has at least one serial entrepreneur.
Founded date	2018

Project Description

Competitive advantage	Accumulated many pilot consumers, 1st mover, Great product design
Traction	Previous Monthly Revenue: $9K, Annual Revenue Growth Rate: 42%

Additional Information

Company Category	B2C
Number of Employees	10-20
Target Market	Domestic Market
Mission	For profit
Location	U.S.
Number of Existing Investors	3 or more

*Assume that all the hypothetical startups work in the industry (or industries) and stage(s) of your interest.

Figure B6: Experiment A Randomly Generated Startup Profile

1. Imagine that Jeffrey Chen and David Zheng's team is guaranteed to accept your investment offer. Compared with firms you have previously invested in, which percentile do you feel this startup belongs to considering its quality?

Extremely Low Quality										Extremely High Quality
0	10	20	30	40	50	60	70	80	90	100

Probability of Generating Higher Return (**Drag the bar**)

2. Considering the potential network and negotiation power of Jeffrey Chen and David Zheng's startup team, what's the probability that this startup team will accept your investment offer rather than that of another investor (Angel, VC, Loans, etc)?

Guaranteed Rejection										Guaranteed Acceptance
0	10	20	30	40	50	60	70	80	90	100

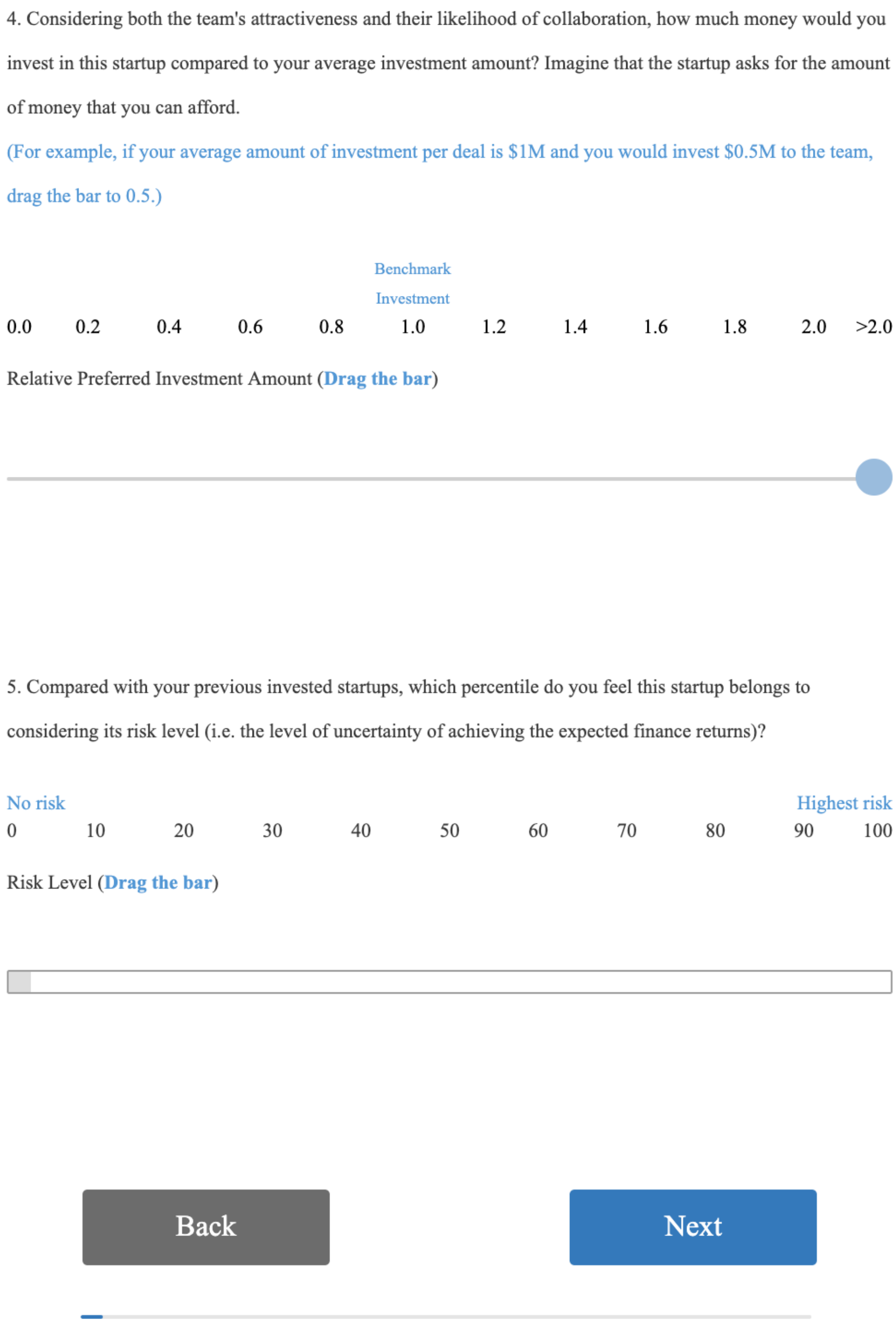
Probability of Accepting Your Offer (**Drag the bar**)

3. If you consider both the team's attractiveness and their likelihood of collaboration, how likely would you be to ask for their contact information or pitch deck?

Will Not Ask										Will Ask
0	10	20	30	40	50	60	70	80	90	100

Probability of Asking for More Information (**Drag the bar**)

Figure B7: Experiment A Evaluation Questions (Part 1)



4. Considering both the team's attractiveness and their likelihood of collaboration, how much money would you invest in this startup compared to your average investment amount? Imagine that the startup asks for the amount of money that you can afford.

(For example, if your average amount of investment per deal is $1M and you would invest $0.5M to the team, drag the bar to 0.5.)

Benchmark Investment

0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 >2.0

Relative Preferred Investment Amount (**Drag the bar**)

5. Compared with your previous invested startups, which percentile do you feel this startup belongs to considering its risk level (i.e. the level of uncertainty of achieving the expected finance returns)?

No risk

Highest risk

0 10 20 30 40 50 60 70 80 90 100

Risk Level (**Drag the bar**)

Back

Next

Figure B8: Experiment A Evaluation Questions (Part 2)

Dear [Investor Name],

Our research team learned about your startup investment experience from Pitchbook and would like to invite you to participate in a research project conducted by the Columbia University Economics Department. Given your expertise in the startup investment, your insight would be indispensable to our research, which we hope would shed light on the entrepreneurial financing process in the U.S. and help the recovery of entrepreneurial activities from recession.

The research project is supervised by Prof. Jack Willis and led by a Columbia Economics Ph.D. student, Ye (Iris) Zhang, who is collaborating with Hash Outliers and the En Lab. The purpose of the project is to understand the current entrepreneurial financing process (for example, investors' preferences for future collaborative startups) and remove the frictions typically found in the fund-raising process using the matching algorithms we have developed. We have developed a matching tool (the "Nano-Search Financing Tool") that can match investors with the best fit startup teams.

Using the tool takes about 20 minutes and involves evaluating 16 hypothetical startup profiles in your invested industry. After evaluating these profiles, the tool uses a newly developed machine-learning algorithm to identify startups who could be a good fit for your investment portfolios from our collaborative incubators. The matched startup teams will try to contact you after 1 month.

Besides the potential investment and collaboration opportunities, we will offer a lucky draw opportunity to thank you for your support of this research project. At the end of July 2020, we will randomly pick 2 survey participants and inform them of the lucky draw results. These 2 participants will be paid in July 2021 according to the startup quality evaluation results they made in the financing tool (that is, the $500 and the extra return based on their quality evaluation results). Details are described on the instruction page and consent form in the matching tool.

To access the tool, please click the link; we have also attached the instruction poster for its use.

Our research team will also use a completely anonymized version of your data to research broader trends in what investors value when investing in startups. We will be glad to share these insights with you when the research is complete.

If you have any questions or would like more detailed information about how the tool will enhance your portfolio construction process, please contact the tool developer and project investigator, Ye (Iris) Zhang (yz2865@columbia.edu).

Thank you very much and have a nice day!

Sincerely,
Ye

--
Ye Zhang
Ph.D. Candidate
Economics Department, Columbia University
Email: yz2865@columbia.edu

Figure B9: Experiment A Recruitment Email (Version 1)

Notes. Version 1 provides both the matching incentive and monetary incentive to randomly selected 11,183 U.S. venture capitalists.

Dear [Investor Name],

Our research team learned about your startup investment experience from Pitchbook and would like to invite you to participate in a research project conducted by the Columbia University Economics Department. Given your expertise in the startup investment, your insight would be indispensable to our research, which we hope would shed light on the entrepreneurial financing process in the U.S. and help the recovery of entrepreneurial activities from recession.

The research project is supervised by Prof. Jack Willis and led by a Columbia Economics Ph.D. student, Ye (Iris) Zhang, who is collaborating with Hash Outliers and the En Lab. The purpose of the project is to understand the current entrepreneurial financing process (for example, investors' preferences for future collaborative startups) and remove the frictions typically found in the fund-raising process using the matching algorithms we have developed. We have developed a matching tool (the "Nano-Search Financing Tool") that can match investors with the best fit startup teams.

Using the tool takes about 20 minutes and involves evaluating 16 hypothetical startup profiles in your invested industry. After evaluating these profiles, the tool uses a newly developed machine-learning algorithm to identify startups who could be a good fit for your investment portfolios from our collaborative incubators. The matched startup teams will try to contact you after 1 month.

To access the tool, please click the link; we have also attached the instruction poster for its use.

Our research team will also use a completely anonymized version of your data to research broader trends in what investors value when investing in startups. We will be glad to share these insights with you when the research is complete.

If you have any questions or would like more detailed information about how the tool will enhance your portfolio construction process, please contact the tool developer and project investigator, Ye (Iris) Zhang (yz2865@columbia.edu).

Thank you very much and have a nice day!

Sincerely,
Ye

--
Ye Zhang
Ph.D. Candidate
Economics Department, Columbia University
Email: yz2865@columbia.edu

Figure B10: Experiment A Recruitment Email (Version 2)

Notes. Version 2 provides only the matching incentive to randomly selected 4,000 U.S. venture capitalists.

Figure B11: Experiment A Recruitment Poster (Version 1)

Notes. Version 1 provides both the matching incentive and monetary incentive to randomly selected 11,183 U.S. venture capitalists.

Figure B12: Experiment A Recruitment Poster (Version 2)

Notes. Version 2 provides only the matching incentive to randomly selected 4,000 U.S. venture capitalists.

Thank you for completing the questionnaire. We will offer you a $15 Amazon Gift Card within 2 days. However, you can also choose to donate a proportion of this $15 to our ${e://Field/racegender} startup club to show your support. Your donation decision is completely anonymous and will not be disclosed to anyone. We will use your donated money to purchase a small gift for one of our ${e://Field/racegender} startup founders.
(For example, if you donate $5 to the club, we will send you a $10 Amazon Gift Card within 2 days and use the donated $5 to purchase a small gift for a ${e://Field/racegender} startup founder in our incubators to give them your anonymous encouragement.)

Please select how much you wish to donate.

$0, no donation

Figure B13: Founder Picture Example in the Donation Section

C Correspondence Test

C.1 Name Generation Process

I generate a list of names that are highly indicative of race (Asian or white) and gender (male or female), combining the approaches of Fryer Jr and Levitt (2004) and Gornall and Strebulaev (2020a). I use the Social Security Administration (SSA) dataset,[98] birth records for selecting first names highly indicative of gender,[99] and 2010 U.S. Census data for generating last names highly indicative of race.[100] The full lists of names are provided in Appendix C.1 Table C1. The following describes the detailed steps for generating these names.

First Names:

Step 1: I started with first names from the Social Security Administration (SSA) dataset of male and female baby names in the U.S. Common names are chosen to mitigate the concern that a distinctively ethnic first name can convey other information besides gender. For example, such confounding information can be social status and economic background of the person (Bertrand and Mullainathan (2004)). Considering that the naming pattern for Asians and white is very similar (Fryer Jr and Levitt (2004)), I do not select indicative first names within an ethnic group.

Step 2: To avoid gender ambiguity, I do the following additional checks. First, I remove ambiguous names, which are defined as names that were in both the top 1,000 male and top 1,000 female lists with a difference in frequency of less than 200,000 times.[101] Then I pick the most frequent 100 names for each gender for further checks.[102]

Second, to remove names that might be perceived as Hispanic or Jewish, we manually checked each potential candidate name and its origin, keeping all the popular Christian names and removing names whose origin is mainly Jewish (countries like Spain, Portugal, or Israel).[103] I further remove names that are strongly indicative of religion (such as Moshe).

Last Names:
I follow exactly the method of Gornall and Strebulaev (2020a) by starting with the most common 1,000 last names in the 2010 U.S. Census data. The white-sounding last names are the 50 most common last names that are more than 85% white and less than 3% Hispanic. The Asian-sounding last names are all 26 last names on the most common list that are more than 85% Asian. I delete the surnames which do not show up in venture capital investors' names. For each selected last name, I search the key word "last name venture capital investor" or "last name angel investor" on Google and LinkedIn. If there is no investor which shows up with this last name, I delete it from the name list. I also remove certain very religious last names. This removed some last names like "Kaur, Vang".[104]

Asian Americans and white Americans have similar first name naming patterns as documented by Fryer Jr and Levitt (2004). Therefore, I have decided to use the last name to indicate the ethnicity status of each created

[98] The SSA dataset is available at https://www.ssa.gov/OACT/babynames/limits.html, accessed on July 27, 2019.

[99] Birth Statistical Master File: https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm, accessed on July 27, 2019.

[100] 2010 Census surnames product: https://www.census.gov/topics/population/genealogy/data/2010_surnames.html accessed on July 27, 2019.

[101] From the histogram of the frequency, we see the majority (74%) of the difference is lower than this number; to be conservative, I choose 200,000 to avoid gender ambiguity.

[102] An alternative method is to construct an index for each name of how distinctively the name was associated with a particular race and gender following Fryer Jr and Levitt (2004). A female name index (FNI) is constructed in the whole sample from SSA data and defined as follows.

$$FNI_{name,t} = \frac{Pr(name|Female,t)}{Pr(name|Female,t) + Pr(name|Male,t)} * 100$$

For selecting female names, I set the cutoff as 99 and keep all the names whose FNI is greater than 99. Among these names, I choose the most frequently used 100 names for female. For selecting male names, I set the cutoff as 3 and keep all the names whose FNI is less than 3. Among these names, I choose the most frequently used 100 names for male. I choose asymmetric cutoffs for female and male FNI due to the fact that the number of male names in the U.S. are much fewer than the number of female names. This method balances the name popularity and also gender unambiguity.

[103] Gornall and Strebulaev (2019) use the name list published by Jorg Michael and removed names that were gender ambiguous in the United Kingdom and as popular in Spain, Portugal, or Israel as in the United Kingdom. We do not feel popular names in these countries are necessarily religious and considering the size of our potential names, manually checking them is feasible here.

[104] An alternative method is to construct a white name index (WNI) and an Asian name index (ANI) following Fryer Jr and Levitt (2004),

fictitious startup founders. To prevent used names from signaling extra information such as a founder's social status, I only select commonly used names that do not have any systematic association with founder' social background.

Additional Check:
I also hire 107 Amazon Mechanical Turk users in the U.S. to confirm that the perception of gender and race elicited by these names was in line with demographic data. For both first names and last names, I exclude any names that are not correctly classified more than 90% of the time. If the number of remaining first names and last names are less than 50 each, I duplicate the process to add names to the waiting list.

After generating names indicative of gender and race each, I randomly paire first names and last names to generate a list of full names assuming that last names do not convey information about gender. I select 50 names for each race-gender combination for randomization. Names of hypothetical female startup founders are shown in Table C1; names of hypothetical male startup founders are shown in Table C2.

To prevent the generated founder names from being associated with famous founder names, I searched LinkedIn to ensure that there were no real famous founders or investors who have the same name and match the key details in the profile. If a conflict is found, I delete the full name and add a new name from the waiting list.

Gender and race are randomized independently. The corresponding names used for each hypothetical startup for both rounds of the correspondence test are provided in Table C3.

C.2 Emailing Process and Preparation Work

Emailing Process—I mainly implement the following two steps to solve the technical difficulties of sending a large number of cold call emails to investors' email inboxes and to passing the existing spam filters.[105] First, before sending large-scale pitch emails in 03/2020, I sent out a testing email (see Figure C1 in Appendix C) which introduces public information about COVID-19 in 02/2020. The testing email is meant to identify which email addresses are invalid and to check the opening rate of cold emails irrelevant to investment opportunities.[106] The opening rate of the testing email after 2 weeks was 2.8%, while the average opening rate of the investment-related pitch emails in this experiment is 11.8%. This indicates that investors only open the emails that they are interested in based on the email subject line and senders.

Second, I used Mailgun's Managed Service,[107] a third-party commercial email API delivery service provider, for sending the large number of emails. Compared with the traditional method of using multiple web hosts to combat spam policies, Mailgun is designed for developers and businesses, with an extremely powerful functionality to track the

---

which is defined as follows.

$$
\begin{aligned}
WNI_{surname,t} &= \frac{Pr(surname|White,t)}{Pr(surname|White,t)+Pr(surname|Non-White,t)} * 100 \\
Pr(surname|White) &= \frac{Pr(surname,white)}{Pr(white)} = \frac{Pr(white|surname) \times Pr(surname)}{Pr(white)} \\
ANI_{surname,t} &= \frac{Pr(surname|Asian,t)}{Pr(surname|Asian,t)+Pr(surname|Non-Asian,t)} * 100
\end{aligned}
$$

I implement similar checks for first names and require that the last name make up at least 0.1% of that race's population, to ensure that last names are sufficiently common.

[105] Different email providers usually use different spam filtering algorithms. However, there are some common patterns for detecting spam emails. First, if there are many invalid email addresses sent out from the same domain at an extremely high frequency (for example, 10 emails sent out per second), then the emails sent are more likely to be labeled as spam. To avoid this, it is helpful for researchers to send a safe testing email identifying the invalid email addresses and then to remove them in the formal recruitment process. Second, if the email contains unverified website links or common words used in spam emails like "Dear," these emails are likely to fail the spam filter. Hence, it is important to use a spam filter testing service to double check the email's contents. However, none of these spam filtering algorithms are correlated with email senders' gender and race.

[106] Invalid email addresses are those that no longer exist or are no longer frequently checked by investors based on the bounced back email notifications. The investor database was constructed between 04/2018-12/2019. Therefore, more than 20% of the collected email addresses are no longer valid due to the high turnover rate.

[107] https://www.mailgun.com/ Mailgun has more than 150,000 customers in 2020. It was founded in 2010 and was a part of the Y Combinator Winter 2011 cohort.

status of each email sent and achieve a high delivery rate through its emailing infrastructure. It also provides developers with complete freedom to customize email sender names, setting the back-end database structure and developing new email tracking functionalities with a user-friendly price compared with Gsuite,[108] which is an email provider from Google. Before automatically sending pitch emails, I used GlockApp, a spam filter testing service provider, to test and improve my pitch email templates.

Following the two-step email sending procedures mentioned above, the response rate is very stable along the whole recruitment process. Gornall and Strebulaev (2020a) used standard methods of sending out a large number of cold call pitch emails and the email response rate declined from 9.0% for the first 4,000 emails to 5.3% for the last 4,000 emails. This situation did not occur in this experiment. Moreover, the email sending procedures in this experiment allow for monitoring multiple investors' information acquisition behaviors without hurting the email delivery rate too much.

Preparation Work — To make sure that the i.i.d assumption holds for the experiment,[109] the preparation work for this experiment is implemented in the following steps. First, to increase the response rate, I match investors with pitched startup ideas based on their industry/vehicle preferences so that,[110] for instance, healthcare-related pitch emails are sent to investors who are interested in the healthcare industry. Second, considering the potential spillover effect within each VC fund,[111] investors receiving the same pitch email ideas come from different VC funds. Each startup pitch email is sent to roughly 1000 investors who all work in different funds. Among these 1000 investors, they are randomly divided into 16 groups because based on the factorial experimental design,[112] founder's gender, race, education and project advantages should be randomized independently. Hence, we have $2 \times 2 \times 2 \times 2 = 16$ groups. Third, it usually takes more than 2 weeks for us to send two sequential pitch emails to the same investor to avoid unnecessary attention and keep the i.i.d. assumption.[113] Each investor received 3 to 5 pitch emails between 03/2020-09/2020.

[108] If researchers have abundant research funding, they can also create multiple Gsuite accounts to combat spam policies. Gsuite is a "company-version" of gmail and is user-friendly to people without strong coding skills. The only drawback is its relatively expensive price, costing $6 per account per month starting in 2020.

[109] Abbreviation for "independent and identically distributed".

[110] For investors recorded in the Pitchbook Database, I use the recorded industry preference for the matching purpose. For investors from other databases, I manually collected their industry preferences from information on their company websites, LinkedIn, and CBInsight. If the manually collected industry information is not accurate, this will increase the noise of the experiment's results and reduce the email response rate. However, it does not affect investors' email opening behaviors.

[111] For some VC funds, they usually have a weekly meeting to discuss promising investment opportunities before replying to cold call pitch emails. If investors receiving the same startup idea come from the same fund, their responses are likely to be correlated. However, this situation will not affect email opening behaviors and email reading time when they just receive pitch emails.

[112] This randomization that the number of treatment group observations is equal to the control group size is mainly to increase the experiment's power.

[113] Gornall and Strebulaev (2020a) waited at least five days to send a sequential email, which raises the attention of some investors who draw attention to these cold emails on twitter in the middle of the experiment. Their experiment was finished between 11/2018-12/2018. To avoid such a situation, I slowed down the pace of sending cold emails and extended the experiment's implementation period.

Table C1: Experiment B First Names Populating Profile Tool

Panel A: Female

Jennifer	Elizabeth	Lisa	Laura	Megan	Emily	Erica
Natalie	Jacqueline	Victoria	Melanie	Tina	Kayla	Kristy
Melinda	Linda	Theresa	Kara	Amanda	Sarah	Amy
Angela	Christina	Rebecca	Tiffany	Mary	Brittany	Samantha
Katherine	Alicia	Monica	Kathryn	Patricia	Anna	Catherine
Veronica	Kathleen	Sandra	Cassandra	Valerie	Amber	Teresa
Allison	Amber	Katrina	Jenna	Megan	Jessica	Melissa
Nicole	Sara	Julie	Christine	Tara	Katie	
(Extra)						
Abigail	Danielle	Michelle	Rachael	Brenda	Margaret	Amanada
Hayley	Madeline	Molly	Vanessa	Rachael	Grace	Heather
Cynthia	Caroline	Karen				

Panel B: Male

Robert	Brian	Kevin	Steven	Thomas	Adam	Patrick
Bryan	Keith	Donald	Peter	Jared	Phillip	Jeffery
Victor	Seth	Alan	Matt	David	Jason	John
William	Andrew	Justin	Anthony	Jonathan	Timothy	Nicholas
Jeremy	Richard	Jeffrey	Benjamin	Paul	Stephen	Nathan
Jacob	Gregory	Travis	Kenneth	Samuel	Edward	Derek
Ronald	Joel	Frank	Dennis	Erik	Philip	Christopher
James	Mark	Scott	Dustin	Zachary	Marcus	Gary
(Extra)						
Vincent	Jack	Luke	Michael	Evan	Joseph	Eric
Shane	Sean	Matthew	Ian	George	Trevor	Charles

Notes. All listed first names which are indicative of gender are used for both the correspondence test experiment and also the lab-in-field experiment. For the correspondence test, these names are used to create fictitious startup founder's names. For the lab-in-field experiment, these names serve as the hypothetical names of startup founders. It covers the popular first names of people who are between 24 years old and 45 years old. To make sure all the names are only indicative of gender, I hire 107 Amazon Mechanical Turks to classify potential names into different genders and provide their feedback on whether these names remind them of other information besides gender (e.g. economic background, race, immigration status, etc). For the all the selected names listed above, more than 98% of Amazon Mechanical Turks correctly classify the names into the corresponding gender. I also delete the names which are indicative of other information. For example, "Chelsea" was deleted because some M-turks feel it is associated with the upper-class; "Luis," "Carlos," or "Antonio" are deleted because they are perceived as more likely to be Hispanic. I also add the first names and last names used in Gornall and Strebulaev (2020a) in the "extra" part.

Table C2: Experiment B Last Names Populating Profile Tool

Panel A: Asian				
Yu	Zhao	Zhang	Jiang	Hwang
Huynh	Luong	Cheung	Hsu	Liang
Li	Hu	Xu	Zhu	Huang
Yang	Kwon	Choi	Nguyen	Pham
Hoang	Luu	Liu	Lu	Chen
Lin	Chang	Chung	Zheng	Xiong
Zhou	Ngo	Truong	Wu	Duong
Cho	Cheng	Yi	Dinh	Tang
Wong	Chan	Ho	Thao	Tsai
Le	Yoon	Wang		
Panel B: White				
Nelson	Russell	Roberts	Rogers	Adams
Cooper	Wright	Cox	Kelly	Phillips
Bennett	Bailey	Collins	Thompson	Stewart
Parker	Evans	Allen	Martin	Anderson
Clark	Campbell	Morris	Reed	Wilson
White	Taylor	Sullivan	Myers	Peterson
Murphy	Fisher	Cook	Hughes	Price
Gray	Moore	Hill	Baker	Hall
Smith	Miller	Ward		
(Extra)				
Hansen	Welch	Hoffman	Meyer	Schmidt
Burke	Beck	Walsh	Carpenter	Schultz
Jensen	Keller	Snyder	Stone	Cohen
Barker	Becker	Schwartz	Larson	Weaver
Carroll				

Notes. The table contains selected last names indicating ethnic identity for hypothetical startup founders. I first create a list of candidate last names combining the results from Method I and the last name list from Kessler et al. (2019). To make sure all the names are only indicative of race and perceived correctly by people, I further hire 107 Amazon Mechanical Turks to classify potential names into different races and provide their feedback on whether these names remind them of other information besides race (e.g. economic background, immigration status, etc.). For the all the selected last names listed above, more than 95% of the Amazon Mechanical Turks correctly classify the Asian last names into the corresponding race and more than 92% of the Amazon Mechanical Turks correctly classified the white last names. I then delete all the ambiguous last names. For example, "Shah" is deleted because many M-turks feel it can also be a middle-eastern name; "Patel" is deleted because they feel it is an Indian name and may not be perceived as a typical Asian name; "Long" is deleted because it can serve as both a white and Asian name. I also delete last names that are related to religion or very rare in the venture capital industry, like "Kaur" and "Vang." I also add the first names and last names used in Gornall and Strebulaev (2020a) in the "extra" part.

Table C3: Experiment B Design, Startup and Entrepreneur Names Used

Panel A: the 1^{st} round

Startup Names	White Female	Asian Female	White Male	Asian Male
VoiceFocus	Kathleen Jensen	Kathleen Yi	Joseph Adams	Kevin Truong
Light Run	Lisa Thompson	Stephanie Lu	Vincent Snyder	Jeffrey Luong
Instrument Tell	Molly Weaver	Jennifer Dinh	Sean Miller	Justin Huang
Sign Reader	Megan Schwartz	Valerie Yu	Evan Meyer	Shane Chan
Bross	Catherine Welch	Rachael Pham	Eric Burke	Ryan Le
Chicky	Rachael Smith	Vanessa Zhu	Robert Reed	Trevor Thao
LoopuDeck	Mary Meyer	Melissa Liu	George Price	Vincent Xu
EasySample	Melissa Larson	Catherine Yang	Matthew Russell	Ian Zheng
YouTubys	Grace Clark	Christine Tang	Justin Hansen	Bryan Hu
OSS	Veronica Russell	Emily Thao	Shane Snyder	Luke Zhao
CPRX	Danielle Cook	Margaret Dinh	Scott Parker	Eric Pham
All-in	Julie Barker	Karen Wong	Marcus Becker	Derek Yoon
SkatED	Kathryn Beck	Abigail Chang	Andrew Moore	George Cheng
GeniusPlot	Christina Parker	Katie Kwon	David Sullivan	Marcus Wang
EasyTry-On	Katherine Snyder	Angela Ho	Richard Cook	Mark Chung
Krysco	Valerie Baker	Amanda Jiang	Patrick Ward	Kevin Hoang
Lens Bioimage Technology	Emily Bennett	Erica Zhou	Adam Hoffman	Peter Cheung
Medprint	Jacqueline Hughes	Patricia Yoon	Ian Cooper	Brian Dinh
BM International	Vanessa Phillips	Mary Luu	Edward Keller	Jack Luu
Vet Technology	Michelle Gray	Natalie Hwang	Jeremy Carroll	Michael Wu
Freight Future	Amanda Meyer	Danielle Cheng	Christopher Cohen	Edward Lin
AfroLab	Madeline Hill	Nicole Xu	Steven Collins	Stephen Liu
SmartTeacher	Jessica Evans	Melanie Ngo	William Welch	Jason Chung
CleanPlanet	Christine Fisher	Megan Liang	Jeffrey Barker	Nicholas Lu
FancyTravel	Melanie Schultz	Rebecca Zhao	Ryan Schwartz	Sean Xiong
MeSafeMicro	Cynthia Keller	Allison Duong	Samuel Kelly	Samuel Ngo
Talently	Caroline Stone	Heather Zhang	Jack Moore	Richard Thao
AgriSoft	Rebecca Miller	Katherine Truong	Gregory Morris	Jonathan Duong
EduPar	Erica White	Caroline Chung	Derek Jensen	Jeremy Jiang
Milkless	Hayley Becker	Christina Hsu	Luke Thompson	William Hwang
Durabuddy	Brenda Bailey	Madeline Tsai	Brian Reed	James Le
Constructech	Samantha Peterson	Samantha Le	Michael Myers	Patrick Nguyen
SolarWat	Patricia Stewart	Brenda Hoang	Thomas Beck	Christopher Huynh

Notes. 33 startups are created for the first round experiment, which was implemented between 03/2020-04/2020. All the startup founders' names are randomly generated using the commonly used first names and last names in the U.S. To prevent the fictitious startup founders from being associated with real people, I search LinkedIn, Google, and available university directories to make sure that no real students from the corresponding universities have the same names. If a conflict is discovered, I replace the conflicting names with other randomly generated names to avoid such a situation. Information of startups used in the later round correspondence test will be updated in the next version of draft.

Table C4: Experiment B Summary Statistics for Hypothetical Startups

Panel A: 1st round

	N	Industry Covered
B2B	13	Media, Music, Fashion, Advertisement, Real Estate, Construction, SAAS, Education, Logistics, Energy, Agriculture
B2C	12	Media, Fashion, Sports, Food, SAAS, Traveling, Pets, Chemical Products, Education
Healthcare	8	Healthcare
Total	33	

Panel B: later round

B2B	13	Entertainment, Media, Packaging, Advertisement, Finance, Management, Education SAAS
B2C	14	Entertainment, Media, Energy, SAAS, Sports, Chemical Products, Food
Healthcare	7	Healthcare
Total	34	

Panel C: Total

B2B	26	Media, Music, Fashion, Advertisement, Real Estate, Construction, SAAS, Education, Logistics, Energy, Agriculture, Entertainment, Packaging, Finance, Management
B2C	26	Media, Fashion, Sports, Food, SAAS, Traveling, Pets, Chemical Products, Education, Entertainment, Energy,
Healthcare	15	Healthcare
Total	67	

Notes. This table reports descriptive statistics for the 67 startups used in the first-round and later-round correspondence tests. All the startups are classified into B2B (Business to Business), B2C (Business to Consumer), and Healthcare following the classification categories of Gornall and Strebulaev (2020a). I also provide more granular industry information about the created startups in the table. Panel A reports the startup category distribution of the first-round correspondence test, which was implemented between 03/2020 and 04/2020 during the outbreak of COVID-19. During the "Chinese Virus" period between 03/18/2020 to 03/24/2020, the three pitch emails sent out include an AI logistics startup (B2B), a healthcare startup, and a startup developing a financial management platform targeting U.S. schools (B2B). The current version of the paper draft only provides the first-round experiment's results. Panel B reports the startup category distribution of the second-round correspondence test, which was implemented in 10/2020 when the economy began to reopen. Panel C reports the startup category distribution of all 67 startups used in the two rounds of correspondence tests. If a startup belongs to both B2B and B2C, I have labeled it as "B2B." In the first round experiment, there were 2 startups belonging to both B2B and B2C. In the second round experiment, there were 3 startups belonging to both B2B and B2C.

Table C5: Experiment B Design, Trace Investors' Email Behaviors

Email Behaviors	Behavior Tracking Mechanisms	Merits	Limitations	Literature
1. Email Opening Rate (time stamp)	Write each pitch email using HTML with a unique ID and insert an one-pixel invisible transparent picture into the email. If the picture is downloaded from the server, I assume the investor opened the pitch email when the picture was downloaded	Increases the experiment's power (high opening rate); only affected by the email's subject line rather than the email's contents	Noisy measurements (Some remote servers prevent users from downloading a picture while others automatically download a picture for their users. However, such server properties are unrelated to the experimental treatment.)	
2. Email Reading Time (time stamp)	Write each pitch email using HTML with a unique ID and insert a large invisible transparent picture (i.e. 500 MB) into the email. Set the speed of downloading the picture from our server to 10KB/s. If only 200KB is downloaded from the server, then the email staying time is 20s.	A continuous variable which measures attention; Increases the experiment's power;	Noisy measurements (Researchers cannot observe directly whether inventors are reading the email or simply leaving the email open while having lunch.)	
3. Multiple Email Opening Rate	If the one-pixel transparent picture inserted in the pitch email is downloaded multiple times as recorded in the server, then I assume the email is opened multiple times. This happens if the same investor opens the email multiple times or the email is forwarded to others who open it later.	Increases the experiment's power; a stronger indicator of investors' interest	Noisy measurement. Researchers cannot differentiate whether the email is opened multiple times by the same investor, or the email is forwarded to others.	
4. Sentimental Analysis of Email Replies	Use LIWC to analyze the sentiment of the content of each email reply. I used the following website which automatically generates analyzed results: http://liwc.wpengine.com/	Relatively objective measurement of the investors' attitudes towards each pitch emails	Low response rate during the recession, hence low experimental power	Hong and Liskovich (2015)
5. Website Click Rate	The Mailgun platform developed this function, and researchers can use it directly. Click here for mechanism explanations provided by Mailgun.	Can be used when investors do not reply to the email	Low website click rate in the entrepreneurial financing setting	Bartoš et al. (2016); Bernstein et al. (2017)
6. Email Response Rate & Reply's Contents	Collected directly from the inbox and spam box	Commonly used callback measurements	Low response rate; The reply's contents may not represent true interest if investors try to be politically correct.	Gornall and Strebulaev (2020a), etc.

Notes. This table provides detailed mechanisms of recording different email behaviors, the merits and limitations of each tracked behavior measurements, and the previous correspondence tests in the literature that used similar participants' behaviors. To realize these functions, I used the Mailgun platform, which is a professionally designed platform for large email campaign activities founded in 2010.

Table C6: Experiment B Heteroscedastic Probit Estimates for Opening Rate by Gender and Race

	Dependent Variable: 1(*Opened*)		
	(1)	(2)	(3) After 03/24
Panel A. Probit estimates			
Female Founder (marginal)	0.010***		
	(0.004)		
Asian Founder		0.006	0.007*
		(0.004)	(0.004)
Panel B. Heteroscedastic probit estimates			
Female Founder (marginal)	0.009***		
	(0.004)		
Asian Founder		0.006	0.008*
		(0.004)	(0.004)
Standard deviation of unobservables, female/male	0.81		
Standard deviation of unobservables, Asian/white		1.12	1.09
Test: ratio of standard deviations = 1 (p-value)	0.27	0.55	0.701
Female-level (marginal)	0.059	-0.021	-0.012
Female-variance (marginal)	-0.050	0.027	0.020
Observations	30,909	30,909	25,525

Notes. This table reports regression results from the heteroscedastic probit estimates for opening rate by gender (female vs. male) after correcting for potential biases from the difference in variance of unobservables. The dependent variable is a dummy variable, which is one when an investor opens the pitch email, and zero otherwise. Marginal effects are computed as the change in the probability associated with being a "female" founder using the continuous approximation, evaluating other variables at their means; the continuous approximation yields an unambiguous decomposition of the heteroscedastic probit estimates. Columns (1) and (2) use all the observations obtained in the first wave. Column (3) uses the observations from pitch emails sent after 03/24/2020. Standard errors are in parentheses. p-values are based on Wald tests. *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance at 1%, 5%, and 10% levels, respectively.

Table C7: Experiment B Heterogeneous Effect of Investors' Response (ESG)

	Dependent Variable: 1(*Opened*)					
	(1) Full	(2) Impact Fund	(3) Common Fund	(4) Full	(5) Impact Funds	(6) Common Fund
Female Founder=1	0.012***	0.103***	0.011***			
	(0.004)	(0.033)	(0.004)			
Asian Founder=1				0.004	0.008	0.004
				(0.004)	(0.032)	(0.004)
Impact Fund=1	-0.048**			-0.010		
	(0.020)			(0.024)		
Female Founder=1 × Impact Fund=1	0.083**					
	(0.033)					
Asian Founder=1 × Impact Fund=1				0.011		
				(0.032)		
US Investor=1	-0.018***	-0.074	-0.017**	-0.018***	-0.080*	-0.017**
	(0.006)	(0.046)	(0.007)	(0.006)	(0.047)	(0.007)
Female Investor=1	-0.015***	-0.057	-0.014**	-0.015***	-0.068*	-0.014**
	(0.006)	(0.039)	(0.006)	(0.006)	(0.040)	(0.006)
Constant	0.197***	0.275**	0.194***	0.202***	0.355**	0.198***
	(0.019)	(0.135)	(0.020)	(0.020)	(0.146)	(0.020)
Startup FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,649	368	23,281	23,649	368	23,281
R-squared	0.006	0.075	0.006	0.006	0.049	0.005

Notes. This table reports the heterogeneous effect of global investors' email opening behaviors in response to randomized pitch emails based on their investment philosophies in the correspondence test. I only include investors whose investment philosophy is available on Pitchbook, which accounts for 76.5% of all the observations. The dependent variable is a dummy variable, which is one when an investor opens the pitch email, and zero otherwise. "Female Founder = 1" is an indicator variable that equals one if the first name of the email sender is a female name, and zero otherwise. Similarly, "Asian Founder =1" is an indicator variable that equals one if the last name of the email sender is an East Asian name, and zero otherwise. "Impact Fund=1" is an indicator variable that equals one if the investor works in a fund with ESG related investment preferences based on Pitchbook Data, and zero otherwise. Such preferences include supporting minority founders, caring about the environmental and social impact, etc. "US Investor=1" and "Female Investor=1" are indicator variables for being a U.S. investor and being a female investor. Columns (1) and (4) reported the regression results for all observations with available investment philosophies. Columns (2) and (5) reported the regression results for investors working in impact funds. Columns (3) and (6) reported the regression results for investors working in common VC funds which do not pursue impact investing strategies. R^2 is the adjusted R^2 for OLS regressions. Standard errors are in parentheses and are clustered at the investor level. *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance at 1%, 5%, and 10% levels, respectively.

Table C8: Experiment B Robustness Check of Heterogeneous Effect of Investors' Response (ESG)

	Dependent Variable: 1(*Opened*)					
	(1) Full	(2) Impact Fund	(3) Common Fund	(4) Full	(5) Impact Funds	(6) Common Fund
Female Founder=1	0.009**	0.023*	0.009**			
	(0.004)	(0.013)	(0.004)			
Asian Founder=1				0.008	-0.022	0.008
				(0.005)	(0.019)	(0.005)
Impact Fund=1	0.014			0.039***		
	(0.010)			(0.015)		
Female Founder=1 × Impact Fund=1	0.014					
	(0.014)					
Asian Founder=1 × Impact Fund=1				-0.029		
				(0.020)		
US Investor=1	-0.015**	-0.044**	-0.010	-0.027***	-0.056**	-0.022***
	(0.006)	(0.018)	(0.007)	(0.008)	(0.023)	(0.008)
Female Investor=1	-0.021***	-0.030*	-0.019***	-0.015**	-0.017	-0.015**
	(0.005)	(0.016)	(0.005)	(0.006)	(0.020)	(0.006)
Constant	0.190***	0.237***	0.184***	0.143***	0.144**	0.146***
	(0.019)	(0.054)	(0.020)	(0.019)	(0.061)	(0.019)
Startup FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,909	2,895	28,014	14,348	1,335	13,013
R-squared	0.006	0.014	0.006	0.006	0.012	0.006

Notes. This table reports the heterogeneous effect of global investors' email opening behaviors based on their investment philosophies in the correspondence test. The definition of impact funds is more general, including both non-profit funds and funds whose description contains suggestive keywords. The dependent variable is a dummy variable, which is one when an investor opens the pitch email, and zero otherwise. "Female Founder = 1" is an indicator variable that equals one if the first name of the email sender is a female name, and zero otherwise. Similarly, "Asian Founder =1" is an indicator variable that equals one if the last name of the email sender is an East Asian name, and zero otherwise. "Impact Fund=1" is an indicator variable that equals one if the investor works in a fund with ESG-related investment preferences based on Pitchbook Data, and zero otherwise. Such preferences include supporting minority founders, caring about the environmental and social impact, etc. "US Investor =1" and "Female Investor = 1" are indicator variables for being a U.S. investor and being a female investor. Columns (1) and (4) report the regression results for all observations with available investment philosophies. Columns (2) and (5) report the regression results for investors working in impact funds. Columns (3) and (6) report the regression results for investors working in common VC funds which do not pursue impact investing strategies. R^2 is the adjusted R^2 for OLS regressions. Standard errors are in parentheses and are clustered at the investor level. *** p<0.01, ** p<0.05, * p<0.1

Table C9: Bias Mechanisms Predicted by Theories (Gender)

Mechanisms	Experiment A	Theory Prediction	Experiment B	Theory Prediction
1. Belief-Based Mechanisms				
1.1 Expected financial return (first moment)	$\surd$ (against)	Q1, $\beta_{female} \neq 0$	$\varnothing$	$\beta_3 < 0$
1.2 Expected variation (second moment)	$\times$	Q5, $\beta_{female} \neq 0$	$\times$	$\sigma^{II}_{FR} \neq 1$
1.3 Strategic channel	$\times$	Q2, $\beta_{female} \neq 0$	$\times$	$\beta_2 < 0$
2. Taste-Based Mechanisms			$\surd$	$\beta_1 > 0, \beta_2 > 0, \beta_3 >= 0$
2.1 Friendly Support	$\surd$	Female investors donate more to female founders.	$\surd$	β_1 is larger for impact funds
2.2 Social Image	$\varnothing$-$\surd$	Male investors donate less to female founders in the donation game	$\varnothing$	
2.3 Others (i.e. Sexual harassment)	$\varnothing$	Older female founders are less likely to be contacted.	$\varnothing$	
Amplifying Mechanisms				
a. Attention Discrimination	$\varnothing$		$\surd$	$\beta_1 > 0$ when Y_{ij} is opening rates and email staying time.
b. Implicit Bias	$\surd$ (against)	Test interaction term of Female Founders and the Second Half Study	$\varnothing$	
3. Other Specific Mechanisms				
3.1 Uninformative Email Behaviors	N/A		$\times$	$\beta_1 > 0$ when Y_{ij} is opening rates and email staying time.
3.2 Fishy Emails	N/A		$\times$	$\beta_2 < 0$

Notes. This table shows the mechanisms predicted by different gender discrimination theories and whether such mechanisms are supported by the empirical results from the correspondence test and the lab-in-field experiment or not. "$\surd$" means such a mechanism is supported by the empirical evidence from the specific experiment. "$\times$" means that such a mechanism is ruled out by the empirical evidence from the specific experiment. "$\varnothing$" means that the experiment does not provide empirical evidence to support or rule out such a mechanism. The parameters used in the correspondence test theory predictions are from the following regressions: $Y_{ij} = \beta_0 + \beta_1 FemaleFounderij + \beta_2 Ivy_{ij} + \beta_3 FemaleFounder_{ij} \times Ivy_{ij} + \alpha_i + \epsilon_{ij}$ with pitch email fixed effect, where Y_{ij} are the behavior measurements like the opening rate dummy, etc. $FemaleFounder_{ij}$ and Ivy_{ij} are indicators of being a female founder and graduating from Ivy League colleges. σ^{II}_{FR} is the ratio of standard errors of female founders' unobservable characteristics and male founders' unobservable characteristics. I found bias towards female founders ($\beta_1 > 0$) in the correspondence test. Hence, all the theory predictions are to explain the reasons why investors prefer female founders. The parameters used in the lab-in-field theory predictions are from the following regressions: $V_{ij} = \beta_0 + \beta_c Characteristics_{ijc} + \alpha_i + \epsilon_{ij}$ with evaluator fixed effect.V_{ij} can be the evaluation of Q1(quality), Q2(collaboration likelihood), Q3(contact), Q4(investment) and Q5(risk). Please note that all the mechanisms can exist at the same time with some mechanisms dominating others in specific experimental settings.

Table C10: Bias Mechanisms Predicted by Theories (Race)

Mechanisms	Experiment A	Theory Prediction	Experiment B	Theory Prediction
1. Belief-Based Mechanisms				
1.1 Expected financial return (first moment)	$\surd$ (against)	Q1, $\beta_{Asian} \neq 0$	$\surd$	$\beta_1 > 0, \beta_2 > 0, \beta_3 < 0$
1.2 Expected variation (second moment)	$\times$	Q5, $\beta_{Asian} \neq 0$	$\times$	$\sigma_{AR}^{II} \neq 1$
1.3 Strategic channel	$\times$	Q2, $\beta_{Asian} \neq 0$	$\times$	$\beta_{Ivy} < 0$
2. Taste-Based Mechanisms				
2.1 Friendly Support	$\surd$	Asian founders receive more donations	$\varnothing$	$\beta_1 > 0, \beta_2 > 0, \beta_3 >= 0$
2.2 Social Image	$\varnothing$		$\varnothing$	
Amplifying Mechanisms				
a. Attention Discrimination	$\varnothing$		$\surd$	$\beta_1 > 0$ when Y_{ij} is opening rates and email staying time.
b. Implicit Bias	$\surd$ (against)	Q1, $\beta_{Asian} \neq 0$	$\varnothing$	
3. Other Specific Mechanisms				
3.1 Uninformative Email Behaviors	N/A		$\times$	$\beta_1 > 0$ when Y_{ij} is opening rates and email staying time.
3.2 Fishy Emails	N/A		$\times$	$\beta_2 < 0$

Notes. This table shows the mechanisms predicted by different racial discrimination theories and whether such mechanisms are supported by the empirical results from the correspondence test and the lab-in-field experiment or not. "$\surd$" means such a mechanism is supported by the empirical evidence from the specific experiment. "$\times$" means that such a mechanism is ruled out by the empirical evidence from the specific experiment. "$\varnothing$" means that the experiment does not provide empirical evidence to support or rule out such a mechanism. The parameters used in the correspondence test theory predictions are from the following regressions: $Y_{ij} = \beta_0 + \beta_1 AsianFounderij + \beta_2 Ivy_{ij} + \beta_3 AsianFounder_{ij} \times Ivy_{ij} + \alpha_i + \epsilon_{ij}$ with pitch email fixed effect, where Y_{ij} are the behavior measurements like the opening rate dummy, etc. $AsianFounder_{ij}$ and Ivy_{ij} are indicators of being an Asian founder and graduating from Ivy League colleges. σ_{AR}^{II} is the ratio of standard errors of Asian founders' unobservable characteristics and white founders' unobservable characteristics. I found bias towards Asian founders ($\beta_1 > 0$) in general in the correspondence test. Hence, all the theory predictions are to explain the reasons why investors prefer Asian founders starting in 04/2020. The parameters used in the lab-in-field theory predictions are from the following regressions: $V_{ij} = \beta_0 + \beta_c Characteristics_{ijc} + \alpha_i + \epsilon_{ij}$ with evaluator fixed effect. V_{ij} can be the evaluation of Q1(quality), Q2(collaboration likelihood), Q3(contact), Q4(investment) and Q5(risk). Please note that all the mechanisms can exist at the same time with some mechanisms dominating others in specific experimental settings.

Subject Line: Monitoring the 2019 Novel Coronavirus (2019-nCoV)
Dear [First Name]:

We are actively monitoring the 2019 Novel Coronavirus (2019-nCoV) and want to share with you important information about the virus's symptoms and current recommendations. The Centers for Disease Control and Prevention (CDC) is working with the World Health Organization as this outbreak, originating in December 2019 in Wuhan City, Hubei Province, China, continues to expand.

Currently, there are few known cases in the U.S. and other countries. However, we want to provide some additional information as this situation evolves. This virus belongs to a family of viruses called "coronavirus." There are other viruses in the coronavirus family that can cause illness in both humans and animals. These viruses can cause either mild illness like a cold or can make people very sick with pneumonia. This particular coronavirus has not been seen previously in humans. There is no vaccine available for this or other coronaviruses.

How is it transmitted?
Since this virus is very new, health authorities continue to carefully watch how it spreads. It is spread from animals to humans and also appears to be spread from person to person. Incubation is likely 5-7 days, but may be up to 14 days.

What are the symptoms?
Fever, cough, and shortness of breath are the most common symptoms. If you have any of these symptoms and have been traveling or in contact with someone that has been traveling in the Asia-Pacific region, please seek medical attention (see below).

Recommendations:

- Please review the CDC Travel Health Notice. The CDC recommends that travelers avoid all nonessential travel to Wuhan, China.
- If you have traveled recently, especially to the Asia-Pacific region, and are experiencing the above symptoms please seek medical attention immediately:
- Wash hands often with soap and water for at least 20 seconds. Use an alcohol-based hand sanitizer, if soap and water is not available.
- Expect additional time at airports and transportation hubs throughout Asia and in major US cities for health screening to prevent spread.

With care for our community,

Confidentiality Disclaimer: This e-mail message and any attachments are private communication and may contain confidential, privileged information meant solely for the intended recipient. If you are not the intended recipient, you are hereby notified that any use, dissemination, distribution or copying of this communication is strictly prohibited. Please notify the sender immediately by replying to this message, then delete the e-mail and any attachments from your system. Thank you.

Figure C1: Experiment B Example of the Testing Email

D Model for Correspondence Test

Assume that the quality (i.e. productivity) of startup depends linearly and additively on two characteristics: X^{I*} which includes standardized observable information in the pitch email; X^{II} which includes unobservable characteristics of each startup. Let G=1 denote being a female founder and G=0 denote being a male founder. (Similar logic can also be applied to Asian founders and white founders.) Define γ as an additional linear additive terms that reflects taste-based bias or belief-based bias (i.e. $E(X_F^{II}) \neq E(X_M^{II})$) based on the founder' gender. Define F as fund-level characteristics, which are normally distributed, independent of X^{II}, and follows the same distribution for female founders and male founders.

D.1 Heckman's Critique

Based on the model from Neumark (2012), the investor would open or reply to an email if a startup's perceived quality exceeds an internal threshold $c'(> 0)$. Then the callback decisions (i.e. the email opening decision or email reply decision) for female and male founders are

$$T(P(X^{I*}, X_F^{II})|G = 1) = 1 \text{ if } \beta_1^{'} X^{I*} + X_F^{II} + \gamma^{'} + F > c'$$
$$T(P(X^{I*}, X_M^{II})|G = 0) = 1 \text{ if } \beta_1^{'} X^{I*} + X_M^{II} + F > c'$$

where X_F^{II} and X_M^{II} are residuals. Assume that X_F^{II} and X_M^{II} are normaly distributed with zero means and standard deviations σ_F^{II} and σ_M^{II}, and the distribution function Φ, then the email opening probabilities are

$$\text{open/reply emails if} X_F^{II}/\sigma_F^{II} > (c' - \beta_1^{'} X^{I*} - \gamma^{'})/\sigma_F^{II} \text{ where } \frac{X_F^{II}}{\sigma_F^{II}} \sim N(0,1)$$

$$(10)\ \underbrace{Pr[T(P(X^{I*}, X_F^{II})|G = 1) = 1]}_{\text{opening/reply probability for female}} = 1 - \Phi[\frac{c' - \beta_1^{'} X^{I*} - \gamma^{'}}{\sigma_F^{II}}] = \Phi[\frac{-c' + \beta_1^{'} X^{I*} + \gamma^{'}}{\sigma_F^{II}}]$$

$$(10')\ \underbrace{Pr[T(P(X^{I*}, X_M^{II})|G = 0) = 1]}_{\text{opening/reply probability for male}} = 1 - \Phi[\frac{c' - \beta_1^{'} X^{I*}}{\sigma_M^{II}}] = \Phi[\frac{-c' + \beta_1^{'} X^{I*}}{\sigma_M^{II}}]$$

Without further assumption on σ_F^{II} and σ_M^{II}, γ is unidentified. The model mentioned above illustrates the Heckman's critique. In a correspondence test, X_F^I=X_M^I = X^I. Consider the situation where $\gamma' = 0$ (no discrimination), but $Var(X_M^I) > Var(X_F^I)$ (i.e. the variance of male founders is larger than the variance of female founders)

Case I: When X^{I*} is low, investors prefer male entrepreneurs (higher variance group) whose $Var(X_M^I)$ is higher. (spurious evidence of discrimination against women)

$$\text{If} \quad \beta_I' X^{I*} < c', \underbrace{\Phi[\frac{-c' + \beta_1^{'} X^{I*} + \gamma^{'}}{\sigma_F^{II}}]}_{\text{super negative}} < \underbrace{\Phi[\frac{-c' + \beta_1^{'} X^{I*}}{\sigma_M^{II}}]}_{\text{not very negative}}$$

Case II: When X^I is high, investors prefer female entrepreneurs (lower variance group) whose $Var(X_F^I)$ is lower.[114] (spurious evidence of discrimination in favor of women)

[114]For example, in Gornall and Strebulaev (2020a), X^{I*} is set as high as possible in order to increase the response rate.

$$\text{If} \quad \beta_I' X^{I*} > c', \underbrace{\Phi[\frac{-c' + \beta_1' X^{I*} + \gamma'}{\sigma_F^{II}}]}_{\text{super positive}} > \underbrace{\Phi[\frac{-c' + \beta_1' X^{I*}}{\sigma_M^{II}}]}_{\text{not very positive}}$$

The HS Critique argument holds for symmetric distributions (Heckman (1998)) and claims that even under ideal conditions, correspondence studies are uninformative about discrimination. The two cases mentioned above show that the relative variances of the unobservables interact with the level of quality set for the pitch email in the correspondence test. Therefore, it is important to check this potential bias from variances of unobservables to avoid spurious evidence of discrimination in favor of women.

Note that the Heckman's Critique comes from the nonlinear binary callback rates used in the correspondence test. It does not apply to the lab-in-field experiment where the outcome variables are continuous and linear.
(i.e. Rating=$\beta_1' X^{I*} + X_F^{II} + \gamma' + F$ for female founders and Rating=$\beta_1' X^{I*} + X_M^{II} + F$ for male founders.)

D.2 Correct Bias Using Neumark Model

Neumark (2012) model shows that when the correspondence test introduces meaning variation of quality that shift investors' response decisions, γ can be identified. The intuition is that when a group has higher variance (i.e. male founders), the effect of its observable characteristics will be smaller. Therefore, checking how quality variation affects investors' callback decisions can help identify the relative variance of the unobservables, and in turn identify γ (i.e. the bias parameter).

The model has the following two assumptions:
a. There are some startup characteristics (i.e. the education background in Experiment 1) in the study that affect perceived quality.
b. β_I is the same for female founders and male founders. (Such assumption cannot be tested in Experiment 1 setting because there is only one significant quality control, which is education background of the startup founder.)

$$\text{Outcome difference} \quad \underbrace{\Phi[\frac{-c' + \beta_1' X^{I*} + \gamma'}{\sigma_F^{II}}]}_{\text{response rate for female founders}} - \underbrace{\Phi[\frac{-c' + \beta_1' X^{I*}}{\sigma_M^{II}}]}_{\text{response rate for male founders}}$$

I can only identify the coefficients relative to the standard deviation of the unobservable, so I normalize the variance. Set $\sigma_M^{II} = 1$ and σ_F^{II} is then the variance of the observable for female founders relative to male founders and $\sigma_{FR}^{II} = \frac{\sigma_F^{II}}{\sigma_M^{II}} = \sigma_F^{II}$ after normalization.

$$\text{Outcome difference} \quad (*) \quad \underbrace{\Phi[\frac{-c' + \beta_1' X^{I*} + \gamma'}{\sigma_{FR}^{II}}]}_{\text{response rate for female founders}} - \underbrace{\Phi[-c' + \beta_1' X^{I*}]}_{\text{response rate for male founders}}$$

(*) can be non-zero due to either (1) $\gamma' \neq 0$ or (2) $\sigma_{FR} \neq 1$, which makes the discrimination not identifiable.

To estimate $\frac{\beta_I}{\sigma_{BR}^{II}}$, β_I, and inferences on their ratio $\sigma_{BR}^{II} = \frac{\sigma_B^{II}}{\sigma_W^{II}}$, I can implement a heteroskedastic probit model which allows the variance of unobservable to vary with gender.

Define i as startup pitch email, define j as investor j. There is a latent variable for perceived quality relative to the threshold, assumed to be generated by

$$T(P_{ij*}) = -c + \beta_I X_{ij}^{I*} + \gamma G_i + \epsilon_{ij}$$

Assume $E(\epsilon_{ij}) = 0$ and var(ϵ_{ij})=$[exp(\mu_\omega G_i)]^2$. μ is also normalized to 0. This model can be estimated via maximum likelihood and the observations are treated as clustered on investor level. Then the estimate of $exp(\omega)$ is equal to σ_{BR}^{II}.

Assume that β_I is the same for female and male in order to identify γ
Observations on male founders identify:$-c$ and β_I
Observations on female founders identify:$\underbrace{\frac{(-c+\gamma)}{exp(\omega)}}_{=\sigma_{FR}^{II}}$ and $\underbrace{\frac{\beta_I}{exp(\omega)}}_{=\sigma_{FR}^{II}}$
The ratio of β_I and $\underbrace{\frac{\beta_I}{exp(\omega)}}_{=\sigma_{FR}^{II}}$ can identify $exp(\omega)$, which is equal to σ_{FR}^{II}.
With c and $exp(\omega)$, the expression of $\underbrace{\frac{(-c+\gamma)}{exp(\omega)}}_{=\sigma_{FR}^{II}}$ identifies γ. If we allow statistical discrimination, which means that $E(X_F^{II}) - E(X_M^{II}) = \mu_{FW}^{II} \neq 0$, then what we identify is $\gamma + \mu_{FW}^{II}$ rather than γ. This is the combination of taste discrimination and the statistical discrimination.

If $\sigma_{FR}^{II} = 1$, then there is no bias from differences in the distribution of unobservabes.
If $\sigma_{FR}^{II} \neq 1$, but we had some evidence on how the level of standardization X^{I*} compares to the relevant startup pitch emails, we could determine the direction of bias.[115]

D.3 Extension of Neumark Model by Adding Strategic Channel

In the Neumark (2012) model described in B.2, the higher the startup perceived quality, the more likely the investor will open this email. However, if some emails are too good ("overqualified"), investors may not want to look at them. Although such mechanism does not play an important role in Experiment 1 setting because better education background positively affects investors' response. Such extra mechanism can be added in the previous model by assuming the following non-monotonic hiring rule:

$$c_2' > \beta_1' X^{I*} + X_F^{II} + \gamma' + F > c_1'$$

Use the following MLE method to estimate the model:

$$\begin{aligned} T_{ij} = & \quad 1\{c_1' < \beta X_1^{I*} + X_2^{II} + \gamma' G + \epsilon_{ij} < c_2'\} \\ T_{ij} = & \quad 1\{(c_1' - X_1^{I*} - \gamma' G)/\sigma_B < X_2^{II} + \epsilon_{ij} < (c_2' - X_1^{I*} - \gamma' G)/\sigma_B\} \\ \prod_{i=1}^{n}(\Phi(\frac{(c_2' - X_1^{I*} - \gamma')}{\sigma_B^F}) - & \quad \Phi(\tfrac{(c_1' - X_1^{I*} - \gamma')}{\sigma_B^F}))^{T_{i\in F,j}=1}(\Phi(\tfrac{(c_2' - X_1^{I*})}{\sigma_B^M}) - \Phi(\tfrac{(c_1' - X_1^{I*})}{\sigma_B^M}))^{T_{i\in M,j}=1} \end{aligned}$$

Such extension is not trivial since it is currently a non-monotonic crossing threshold model and it is hard to non-parametrically estimate such models. (see Lee and Salanié (2018))

[115] Stata Code: dprobit, vce(cluster)

E Proof of "Leave-One-Out" Estimator

The "leave-one-out" estimator developed here can be used to generate the heterogeneous effect based on the evaluator's decision in the IRR experiment. By taking advantage of the new variation within each individual, we can test discrimination channels for the "anti-minority" subgroup and the "minority-friendly" subgroup (defined by whether they prefer contacting or investing in the minority group). Such estimator helps researcher to tell a more detailed story by holding a magnifier.

Proof:
Investor i evaluates the j^{th} randomly generated startup profile. Currently, since we have I investors, each evaluates J profiles, $i \in \{1, 2, ..., I\}, j \in \{1, 2, ..., J\}$, we can run pooled regressions to test group-level preferences.

$$Y_{ij}^{(k)} = X_{ij}\beta_i^{(k)} + \alpha_i + \epsilon_{ij}^{(k)} \tag{2}$$

$Y_{ij}^{(k)}$ means investor i evaluated the k^{th} question for the j^{th} generated profile. $k \in \{1, 2, 3, 4\}$ since each investor needs to provide the answers to Q1 (quality), Q2 (collaboration), Q3 (contact) and Q4 (investment). For simplicity, let's assume X_{ij} contains only one gender indicator.
$X_{ij} = 1$ if the founder's gender is female for the j^{th} generated profile evaluated by investor i.
$X_{ij} = 0$ if the founder's gender is male for the j^{th} generated profile evaluated by investor i.
Due to the experiment design, $\epsilon_{ij}^{(k)} \perp \epsilon_{ij'}^{(k)}$ if $j \neq j'$, however, $\epsilon_{ij}^{(k)} \not\perp \epsilon_{ij}^{(k')}$ if $k \neq k'$
(Note: we need a little bit of structure for the assumption that $\epsilon_{ij}^{(k)} \perp \epsilon_{ij'}^{(k)}$ if $j \neq j'$)

$$\epsilon_{ij}^{(k)} = \eta_i^{(k)} + v_{ij}^{(k)}, v_{ij}^{(k)} i.i.d \tag{3}$$

$\eta_i^{(k)}$ is the fixed effect and will enter the constant term if we run the individual-level regressions. Under this residual structure, we can have the following assumption without loss of generality: $\epsilon_{ij}^{(k)} \perp \epsilon_{ij'}^{(k)}$ if $j \neq j'$). For simplicity, let's classify investors based on $\beta_i^{(3)}$, and define
"anti-minority" investors: $\beta_i^{(3)} < 0$, i.e. investors who do not want to contact the minority founder's startups;
"minority-friendly" investors: $\beta_i^{(3)} > 0$, i.e. investors who prefer contacting the minority founder's startups;

Case i: (Ideal Case) If $\beta_i^{(1)}$ is observable or predetermined (i.e. $\beta_i^{(1)} \perp \epsilon_{ij}^{(k)}$), the classification method is fine. We can divide those 22 investors into 2 groups based on the sign of $\beta_i^{(1)}$, then run the following regression:

$$Y_{ij}^{(k)} = \gamma_1 1(\beta_i^{(1)} < 0)X_{ij} + \gamma_2 1(\beta_i^{(1)} > 0)X_{ij} + \alpha_i + \epsilon_{ij}^{(k)}$$

since $1(\beta_i^{(1)} < 0)X_{ij} \perp \epsilon_{ij}^{(k)}, 1(\beta_i^{(1)} > 0)X_{ij} \perp \epsilon_{ij}^{(k)}$, there is no endogeneity problem.

Case ii: If $\beta_i^{(1)}$ is unobservable, the previous naive classification method (or estimation method) generates biased estimated results.

a. Why? This is a typical "generated regressor problem".

If $\hat{\beta_i^{(1)}} = \frac{\sum_j X_{ij}Y_{ij}^{(1)}}{\sum_j X_{ij}^2} = \beta_i^{(1)} + \frac{\sum_j X_{ij}\epsilon_{ij}^{(1)}}{\sum_j X_{ij}^2}$, then $1(\hat{\beta_i^{(1)}} < 0)X_{ij} = 1(\beta_i^{(1)} + \frac{\sum_j X_{ij}\epsilon_{ij}^{(1)}}{\sum_j X_{ij}^2} < 0)X_{ij}$, which $\not\perp \epsilon_{ij}^{(k)}$ since $\epsilon_{ij}^{(1)} \not\perp \epsilon_{ij}^{(k)}$. Similar problem applies to $1(\hat{\beta_i^{(1)}} > 0)X_{ij}$. Then we have the endogeneity problem ("Y (or the error term) enters the right side of the regression, which is wrong.").
b. To solve this "generated regressor problem", we can use the "leave-one-out" technique widely used in ML. (Thanks to the new variation within each individual, which is unavailable in traditional empirical setting.)
Step 1: for each i & j, estimate $\beta_i^{(1)}$ leaving the j^{th} observation out: $\hat{\beta_{ij}^{L(1)}} = \frac{\sum_{j' \neq j} X_{ij'}Y_{ij'}^{(1)}}{\sum_{j' \neq j} X_{ij'}^2}$ (when

$|J| \to \infty, \beta_{ij}^{\hat{L}(1)} \xrightarrow{p} \beta_i^{(1)}$ for each j). Now we have $I \times J$ estimated $\beta_{ij}^{\hat{L}(1)}$

Step 2: classify $I \times J$ $\beta_{ij}^{\hat{L}(1)}$ into two groups based on their signs. (This means that investor i can enter both the "anti-minority" group and the "minority-friendly" group in a finite sample. However, as $|J| \xrightarrow{p} \infty$, this situation will not happen)

Step 3: run the pooled regressions

$$Y_{ij}^{(k)} = \gamma_1 1(\beta_{ij}^{\hat{L}(1)} < 0)X_{ij} + \gamma_2 1(\beta_{ij}^{\hat{L}(1)} > 0)X_{ij} + \alpha_i + \epsilon_{ij}^{(k)}$$

Now, $\beta_{ij}^{\hat{L}(1)} \perp \epsilon_{ij}^{(k)}$ since $\beta_{ij}^{\hat{L}(1)}$ has left the j^{th} term out (i.e $\epsilon_{ij}^{(1)}$ does not enter $\beta_{ij}^{\hat{L}(1)}$), which breaks the connection with $\epsilon_{ij}^{(k)}$. (Remember our assumption from the experiment design: $\epsilon_{ij}^{(k)} \perp \epsilon_{ij'}^{(k)}$ if $j \neq j'$, however, $\epsilon_{ij}^{(k)} \not\perp \epsilon_{ij}^{(k')}$ if $k \neq k'$), then $1(\beta_{ij}^{\hat{L}(1)} < 0)X_{ij} \perp \epsilon_{ij}^{(k)}$, there is no endogeneity problem using this estimation method.

Note: Theoretically, we can classify the group based on $\beta_i^{(k)}$ for $\forall k$. The interpretation of the results will change since the "anti-minority" group and the "minority-friendly" group are defined by different $\beta_i^{(k)}$ for different k. Depending on the research question, we can choose the most reasonable k

Q.E.D.