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

Ye Zhang †

December 28, 2020

Click here for the latest version

Abstract

This paper examines discrimination based on startup founders' gender, race, and age by early-stage investors, using two randomized controlled trials with real venture capitalists. The first experiment invites U.S. 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 sends hypothetical pitch emails with randomized startups' information to global venture capitalists and compares their email responses by utilizing a new email technology that tracks investors' detailed information acquisition behaviors. I find three main results: (i) Investors are biased towards female, Asian, and older founders of relatively low quality startups; while biased against female, Asian, and older founders of relatively high quality startups. (ii) These two experiments identify multiple coexisting sources of bias. Specifically, statistical discrimination is an important reason for "anti-minority" investors' contact and investment decisions, which was proved by a newly developed consistent decision-based heterogeneous effect estimator. (iii) There was a temporary, stronger bias against Asian founders during the COVID-19 outbreak, which started to fade in April 2020.

Key Words: Venture Capital, Entrepreneurship, Discrimination, Field Experiments

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, Bernard Salenie, 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. They provided extremely valuable feedback on how to improve the experimental methods. 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. Please email the author for code examples.

†Columbia University Economics Department. Email: yz2865@columbia.edu.

1 Introduction

There is a heated debate about whether early-stage investors are biased against female founders, founders of color, and older founders with practitioners, policy makers, and researchers often disagreeing. First, the well-documented, stark funding gap between male-founded start-ups and female-founded start-ups in all stages of the financing process has raised concerns about gender bias (Ewens and Townsend (2020), Guzman and Kacperczyk (2019)) in the venture capital (VC) industry.[1] This concern mainly stems from the fact that about 80% of VC investment professionals are men, and investors may have implicit or unconscious bias against female founders.[2] Second, the documented less favorable treatment received by founders of color during the fundraising process also raises concerns about racial bias (Henderson, Herring, Horton and Thomas (2015)). Based on Gompers and Wang (2017b), 87% of U.S. venture capitalists are white, and investors may also have 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.[3] Third, although older entrepreneurs are a burgeoning growth of innovation force, a large amount of anecdotal evidence and surveys indicate wide-spread ageism in the entrepreneurial community.[4] Such discrimination questions are of critical importance for maintaining social fairness (Fang and Moro (2011)) and assessing the efficiency (Bertrand (2020)) of capital allocation in high-impact startups.

Examining these suspected biases is empirically challenging, mainly due to data limitations and the lack of exogenous variations. Moreover, there are conflicting results in existing literature. Most existing data sources do not observe startups' unique comparative advantages (Ewens and Townsend (2020)).[5] This means that non-experimental studies which do not exploit exogenous variation can suffer from severe omitted variable bias problems, making it difficult to generate causal evidence from them. Gornall and Strebulaev (2020a) make the first attempt to causally test gender and racial discrimination in the VC industry using a correspondence test, which is a standard randomized controlled trial (RCT) method. They compare U.S. venture capitalists' email response rates to fictitious pitch emails with randomized email senders' names and find that, surprisingly, investors reply more frequently to emails sent by female

[1] Gompers and Wang (2017a) 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.

[2] See the NVCA-Deloitte Human Capital Survey Report https://www2.deloitte.com/content/campaigns/us/audit/survey/diversity-venture-capital-human-capital-survey-dashboard.html

[3] "Asians" in this paper primarily stands for "East Asian" groups who have origins mainly from China, Korea, Vietnam, etc. According to Gompers and Wang (2017b), 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-representative minorities is an important question. However, my experimental design needs to adjusted for future researchers to study these important questions.

[4] See Forbes "The Biggest Bias In Tech That No One Talks About" (April 10 , 2019) by Maren Thomas Bannon, an early-stage technology venture capitalist.

[5] Several recent papers have made progress in obtaining nearly "ideal data" that cover all people that want to be entrepreneurs. See Guzman and Kacperczyk (2019), Ewens and Townsend (2020), Hu and Ma (2020), Hebert (2020), and other papers using Census data such as Cen (2020). However, these data still do not include all important startup characteristics, for example, the founder's passion or the project's trade secrets.

and Asian names. This indicates that early-stage investors are biased towards female and Asian founders, while other descriptive papers (Ewens and Townsend (2020), Guzman and Kacperczyk (2019), and Henderson et al. (2015)) show that early-stage investors are biased against female and Asian founders. However, this experiment method suffers from standard limitations of the correspondence test. First, these surprising results in the cold call email setting, which is not the mainstream fundraising method, may not generalize to other fundraising situations. Second, they were unable to introduce meaningful quality variations due to the "low-response-rate" problem. This makes it difficult for the experiment to test the underlying source of discrimination,[6] which is crucial for welfare analysis and policy-making (Bohren, Imas and Rosenberg (2019a), Neumark (2012)). Third, the correspondence test method generally involves deception and only observes investors' initial contact interest, which may be weakly related at best to investment interest or other real economic outcomes.

To establish causality and address the experimental limitations mentioned above, this paper implements the following two complementary RCTs by recruiting real VC investors mainly from the U.S. and other English-speaking areas. I also construct a global, individual-level VC investor database for these two experiments. The first experiment follows the recent RCT method (i.e., lab-in-field experiment) and has a powerful internal validity.[7] It provides stronger incentives for investors to reveal their true investment preferences and tests detailed underlying sources of discrimination. The second experiment follows the standard RCT method (i.e., correspondence test) with an advanced design and has a powerful external validity.[8] It checks how well the results can be generalized among a large number of investors and improves mechanism testing compared to the standard correspondence test design. Combining both methods in that way allows me to paint a nuanced picture about discrimination while also making the methodological contribution of comparing the two methods.

I start with the recent new RCT experimental methodology for testing discrimination - Incentivized Resume Rating - referred to as Experiment A in this paper. To implement this experiment, I work with several accelerators and build a "Nano-Search Financing Tool," which is 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.[9] Investors know the profiles are hypothetical, but they are willing to provide truthful evaluations so that the algorithm works better 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 accurately investors' evaluation results are, the larger the monetary award the lottery winners will receive.[10]

[6] The experiment of Gornall and Strebulaev (2020a) does not introduce variations of startup characteristics that affect the perceived profitability of that startup in their experimental design because of the "low-response-rate" problem. The response rate to their cold call pitch emails is about 6.5% even though all the emails were designed to be as attractive as possible. This "low-response-rate" problem reduces the correspondence test's experimental power, making it difficult to introduce variations in startup quality.

[7] Internal validity measures whether a study establishes a trustworthy cause-and-effect relationship between a treatment and an outcome. It also reflects how powerful an experiment can be to eliminate alternative explanations for a finding.

[8] External validity refers to how generalizable the findings are in other settings. For example, whether results are stable among a larger population or at different times.

[9] Disentangling the potential sources of bias requires researchers to separate various belief-based sources (i.e., "statistical discrimination") (Bertrand and Duflo (2017), Altonji and Blank (1999)) from different taste-based sources (i.e., "animus"). This disentanglement is difficult in the discrimination literature (Gneezy, List and Price (2012)) despite its importance.

[10] Although a "monetary incentive" is noisier than a "matching incentive," it is friendly for researchers without many social connections

This part essentially follows the new incentivized resume rating (IRR) experimental paradigm created by Kessler, Low and Sullivan (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. As in a standard dictator game, investors can accept it or anonymously donate a portion of the $15 to randomly displayed startup teams. Investors are also told that I will use the donated money to purchase small gifts for the corresponding real startup teams in the collaborative accelerators to give founders encouragement and support from the entrepreneurial community during the COVID-19 pandemic. This part essentially follows the representative dictator 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 three main findings. First, although this experiment does not find group-level explicit bias against minority founders (i.e., female, (East) Asian, and older founders), it shows the evidence of implicit bias against female and Asian founders. The investment interest in female founders and the quality evaluations of both female and Asian founders significantly decline when investors are fatigued. Specifically, investors in tech sectors are implicitly biased against female founders because these founders' startups are considered to be less profitable (i.e., statistical discrimination).[11] Similarly, in the "higher contact interest" situations, investors are also implicitly biased against Asian founders. The magnitude of this implicit bias against female and Asian founders is more than 40% of the effect of going to an Ivy League college in investors' evaluations. Second, the distribution effect shows heterogeneity in this bias: investors are biased towards female, Asian, and older founders in the "lower contact interest" or "lower stake" situations (defined as the situations in which investors have less likelihood to contact the team). However, investors are biased against female, Asian, and older founders in the "higher contact interest" or "higher stake" situations (defined as the situations in which investors have more likelihood to contact the team).[12] The preference towards older founders stems from the belief that the older founders impose less risk. This phenomenon reconciles contradictory results in existing literature by demonstrating how the direction of bias depends on context. Third, the donation results show the taste-driven homophily effect that male investors are less likely than female investors to provide support and encouragement to female founders.[13] On average, male investors donate $3 less to female founders compared to similar male founders. However, female investors can donate a little bit more to female founders.

Experiment A shows that some investors have implicit biases against minority founders, while there are also some impact funds that support minority founders a lot. So how divided is the investment community in terms of their attitude towards minority founders, and what separates us? To answer this question, I develop a consistent decision-based heterogeneous effect estimator using the "leave-one-out" technique in Experiment A.[14] This estimator uses exogeneous "within-individual" level randomization to test what the separate driving forces are of the "anti-minority"

and helps increase the experiment's sample size.

[11] Implicit bias refers to the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner.

[12] In Section 5, I provide the effect of the founder's gender, race and age across the distribution of the investor's contact interest.

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

[14] Junlong Feng provides crucial help and discussions for developing this estimator. Our ongoing research work will provide the generalized form of the estimator and guidance on its application in the real world.

groups and the "pro-minority groups," which are defined by investors' 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 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.

Experiment A's design has the following merits and limitations. For the merits, it first provides strong incentives for investors to reveal their true investment preferences, including both their initial contact interest as well as their investment interest.[15] Second, it demonstrates how investors evaluate startups, taking into consideration the whole spectrum of startup quality, from a "low contact interest" situation to a "high contact interest" situation. Third, by using "within-individual" level randomization, this RCT elicits investors' "individual-level" preferences in addition to "group-level" preferences (i.e., group-level average treatment effects),[16] which is crucial to testing implicit bias and "decision-based" heterogeneous effects.[17] However, the limitations of Experiment A include the sample selection bias and the consent form effect. This experiment recruits a relatively small number of investors, who may not be representative of all investors, and it essentially trades some external validity in exchange for stronger internal validity. Moreover, I provide consent forms to all participants and inform them of the research purpose. Therefore, the bias found in Experiment A is likely to be the lower bound of investors' true bias.

Given the limitations of Experiment A, it is important to also implement the second complementary experiment using the standard RCT method and to examine whether results are consistent for a larger sampling pool. Therefore, I follow up with a correspondence test using an advanced design, referred to as Experiment B in this paper. During the COVID-19 outbreak (03-04/2020) and the economy's re-opening (10/2020), 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 background, and randomized startup project characteristics displayed in both the email's subject line and in the email's contents.[18] By utilizing new email tracking technology, I can monitor detailed information acquisition behavior for each investor. In addition to the email response rate used in the standard correspondence test, I also track each investor's email opening behavior, time spent on pitch emails, click rate on the inserted startup's website, and the

[15] Investment activities in the VC industry usually involve multiple stage investments, from seed round to pre-IPO stage. The design of Experiment A also provides the possibility of investigating each detailed investment stage by carefully designing the incentive structure and each evaluation question.

[16] The standard RCT method implements "cross-individual" level randomization, which means randomly selected experimental subjects belong to the control group while another randomly selected group belongs to the treatment group. By comparing the outcome variables of the control and treatment groups, investors can identify group-level average treatment effect

[17] This heterogeneous effect is more and more important for studying a divided society or communities where everyone has their own independent critical thinking and judgements. With the trend of increasing field work in economics, it is exhilarating to see the possibility of future research work extending the current experimental and econometric tools to this new setting and exploring this vigorous research area.

[18] When this research project began at the beginning of 2018, I started with two alternative, more ethical experimental designs. Unfortunately, both of them failed for different reasons and the related discussions of alternative designs are provided in Section 4.1. I choose the current version after long discussions about the experiment's feasibility and risk with Columbia's IRB (i.e., the institutional review board).

contents in email replies. These new experimental designs and behavior 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)). This experimental design follows the correspondence test experiment and essentially sacrifices some internal validity - measuring an imperfect proxy of what researchers actually care about - in exchange for stronger external validity.

Experiment B's results confirm that investors are biased towards female and Asian founders in a "low contact interest" situation (i.e., the pitch email setting).[19] In general, sending pitch emails using female names increases the email opening rate by 1% compared to using male names, and increases the email opening rate by 10% for impact funds when using female names. Similarly, starting in 04/2020, investors spent 10% more time on pitch emails with Asian names, and the opening rate is also 0.7% higher than pitch emails with white names. In addition, revealing that the founding team has an excellent educational background can increase pitch email opening rates by roughly 1%. I also find a temporary, stronger 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, although this bias quickly reversed starting in April, 2020. Compared with Gornall and Strebulaev (2020a), I further test the underlying sources of discrimination. Results show that the bias towards female founders is likely driven by taste-based reasons, while the bias towards Asians is likely driven by belief-based reasons.[20]

Experiment B's design has the following merits compared to the standard correspondence test and also some standard limitations. For the merits, by generating randomized information about a startup in the email's subject line and comparing the email opening rates, it first increases the experimental power and solves the "low-response-rate" problem in the cold call email setting. Second, by tracking detailed information acquisition behaviors of investors and introducing meaningful variation in startup quality, this experiment tests more mechanisms and hence increases its internal validity compared with the standard correspondence test. Third, by implementing this experiment multiple times, it is feasible to check how stable results are along the time dimension. This experimental design helps future researchers to study similar "cherry-picking" markets as well as the labor market even in a recession,[21] when field work usually suffers from the "low-response-rate" problem. However, the limitations of Experiment B are also very obvious. Sending cold call pitch email is not the mainstream fundraising method, and email behaviors can be different from investment behaviors. These limitations are mitigated by Experiment A.

The contribution of this paper is both empirical and methodological. Empirically, it provides the following contributions. First, using the recent RCT method, this paper provides experimental evidence that confirms the existence of investors' implicit bias against female and Asian founders. It also shows that compared to female investors, male investors are less likely to provide anonymous support to female founders. Second, using the standard RCT method with

[19]Sending cold call pitch emails is not the mainstream fundraising method, accounting for less than 12% of total deal flows (Gompers, Gornall, Kaplan and Strebulaev (2020)).

[20]For example, Asian-led startups are perceived to have relatively higher quality by investor in the cold call pitch email setting starting in April.

[21]Gompers et al. (2020) surveyed 885 institutional venture capitalists and document that VCs invest in only 1% of the start-ups they consider. Evaluators can also be very selective in the college admission process, high-skilled job markets, and etc.

an advanced design, this paper documents a temporary, stronger bias against Asian founders during the COVID-19 outbreak and shows how discrimination can be affected by big social events. Third, this paper reconciles contradictory results in the literature by showing how the direction of bias depends on context in the entrepreneurial financing setting. Therefore, this paper empirically contributes to both discrimination literature and entrepreneurial financing literature.

Methodologically, this paper mainly contributes to the field and lab-in-field experiment literature with the following four improvements. First, Experiment A combines the IRR preference elicitation technique and the dictator experiment, allowing it to directly test belief-based discrimination mechanisms and taste-based discrimination mechanisms. Second, the developed decision-based heterogeneous effect estimator using "within-individual" level randomization measures how divided society is and what separates us. Third, the incentive structure provides the possibility of applying the IRR experiment in other settings in addition to a two-sided matching market. Fourth, Experiment B solves the "low-response-rate" problem in the "cherry-picking" market by introducing variations in the email's subject line and tracking investors' new, detailed information acquisition behaviors.

To the best of my knowledge, this is also 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),[22] 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 powerful ability to identify subtle mechanisms, test heterogeneous effects and distributional effects, and generate results about later-stage decisions. Despite these impressive merits, the current version of the IRR experiment is likely to be a good complementary experimental design rather than a full substitute of 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 global VC investor database by merging multiple commercial databases with manually collected data. Section 3 presents the design of Experiment A and analyzes investors' evaluations of startup profiles. Section 4 describes the design of Experiment B and analyzes investors' information acquisition behaviors. Section 5 reconciles the contradictory results from both experiments and the contradictory results in the literature by analyzing the distributional effect. It also discusses the complementarity of these two experiments and the related policy implications. Section 6 studies the decision-based heterogeneous effect to measure how divided the investment community is and what separates us. Section 7 concludes.

[22] 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.

2 Data

I have constructed a cross-sectional, individual-level global venture capitalist database, which contains the most recently updated demographic information and contact information for 17,882 investors before 02/2020. This database contains only investors in English-speaking areas whose email addresses are verified by the testing email used in the correspondence test. Since the experiments are implemented in English, I did not include investors from the Europe and most Asian areas. Therefore, strictly speaking, the database used in this paper is a subset of a more comprehensive global venture capitalist database that also contains investors from Europe and China.

This global database combines the following commercial databases: Pitchbook, ExactData, CB Insight, SDC New Issues Database VentureXpert, and Zdatabase.[23] For investors whose contact information is not available in these commercial databases, I have supplemented this database with contact information collected from RocketReach. 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. Detailed database descriptions and the key variable construction process are provided in 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.[24] 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. The map of investors' global geographical distribution is provided in Figure 1, and the U.S. geographical distribution is provided in Figure 2. 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.[25] Senior investors, who are partners, president, C-level managers, or vice president and above, account for 84% of total investors in our database based on available online

[23] Many of these commercial databases are not free and require researchers to sign a data contract for academic purposes.

[24] 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.

[25] 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.

information. Most investors are institutional investors, and angel investors, who only account for 11% of our sample investors. 61% of investors attended graduate schools and more than 30% of them attended top universities. This is consistent with Gompers and Wang (2017a), who show that VC investors are usually better educated than the average level. Only 2% of all investors work in not-for-profit impact funds.[26] If I use the indicative key words in the fund descriptions to classify the VC funds following Barber, Morse and Yasuda (2020), this percentage increases to 6%-8% depending on the key word selection method.

3 Experiment A: Lab-In-Field Experiment

Experiment A, as a lab-in-field experiment,[27] is designed to elicit investors' investment preferences with a stronger incentive and to solve the limitations of the standard RCT method, like the correspondence test. It combines the following two preference elicitation techniques: the IRR experiment, designed to directly test belief-based discrimination mechanisms, and the dictator experiment, designed to directly test taste-based discrimination mechanisms. I invited real U.S. venture capitalists to try using a "Nano-Search Financing Tool," which is a machine learning, algorithm-based matching tool for investment opportunities. In the first part of the tool (i.e., the IRR experiment), investors need to evaluate multiple randomly generated startup profiles, which they know to be hypothetical, in order to be matched with real, high-quality startups from the collaborative incubators. In the second part of the tool (i.e., the dictator experiment), each investor will receive an unexpected $15 Amazon Gift Card for their participation. Investors can choose whether to keep the $15 or donate a proportion of the $15 to randomly displayed startup teams. The donated money is used to purchase small gifts for real startup teams and provide investors' anonymous support during the pandemic recession.

An experimental setting that develops data-driven methods to help investors evaluate potential deals is not unique in the venture capital industry. A few incubators and VC funds have done extensive work on developing machine learning algorithms to help evaluate investments.[28] However, considering that several important startup characteristics, such as the founder's passion and confidence, cannot be fully quantified by the data, these data-driven methods are usually designed to complement existing, mainstream, person-to-person multiple stage investment strategies rather than to fully substitute for the existing due diligence method.

This section is organized as follows. Section 3.1 introduces the experiment's design and implementation details. Section 3.2 describes the results of the analysis of investors' evaluations and donation decisions. Section 3.3 discusses the robustness test and the limitations used in this experiment.

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

[27] Lab-in-field experiments provide the same clean experimental environment as a lab experiment. However, the subjects are the targeted community in the field, which are real venture capitalists in this paper.

[28] For example, Techstars, Social+ Capital, Citylight Capital, etc. Also, Open Scout, a startup working with the Angel Capital Association (ACA), is designing platforms to connect founders with investors based on shared interests rather than shared network on their platforms.

3.1 Experimental Design

3.1.1 *Investor Characteristics and Recruitment Process*

Experiment A was implemented from 03/2020 - 09/2020 using only online recruitment methods.[29] I sent invitation emails together with the instruction posters to the 15000+ U.S. venture capitalists who also participated in Experiment B (see Appendix B Figure B6 and Figure B7 for the recruitment emails, Figure B8 and Figure B9 for the instruction posters). Both the recruitment emails and posters emphasize the matching purpose of this tool. However, investors were also notified of the research purposes, and they understand that the anonymized data are used for studying investors' preferences for different startups' characteristics as required by IRB. Therefore, this study has the ecological validity of a "natural field experiment," except that the subjects know that their data will also be used for academic research.

There are, in total, 69 real U.S. investors from 68 different funds participating in this project,[30] which provides 1216 startup profile evaluation results.[31] The number of recruited experimental participants is comparable with Kessler et al. (2019), and one 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 advantage is crucial for the experiment to succeed in an environment in which it is hard to recruit large number of subjects.

Similar to the majority of experiments, Experiment A, with roughly a 0.5% response rate,[32] also has sample selection bias during the recruitment process. Based on the observable investor information, Table 2 reports the summary statistics of participants' backgrounds, showing that our sample investors are more likely to come from larger VC funds and to be minority founders.[33] The average asset under management (AUM) of the VC funds is \$547.46 million, which is larger than the average AUM of \$444.44 million in 2019 based on an the NVCA survey. During our 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." 42% of investors in the sample are from minority groups (i.e., Asian, Hispanic, African, etc.), which is higher than the percentage of minority investors in the U.S.[34] However, our sample investors are representative in other dimensions. Recruited investors are mainly early-stage

[29] During the pandemic, Columbia IRB paused all field work which involves person-to-person activities due to COVID-19.

[30] Recruiting real venture capitalists is crucial to understand startup investing strategies because venture capital investment involves very specific skills. Carpenter et al. (2008) documents that lab experiment results provided by college students are very different from results provided by community members, which confirms the importance of lab-in-field experiments. Moreover, the valuation of startups requires relevant high skills (Gornall and Strebulaev (2020b)).

[31] At the beginning of the study, each investor evaluates 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. One investor participated twice for different funds. Results are similar after removing the first 6 investors. As more investors participate in Experiment A, I will update the results in the future.

[32] Future researchers can recruit investors by participating in different real events after the COVID-19 pandemic or collaborating with certain associations (i.e the Angel Capital Association (ACA) or the National Venture Capital Association (NVCA)) to increase the response rate.

[33] Recruited investors are likely to be the investors who are still active during the recession. Based on many investors' email replies, investors usually choose not to participate in this research because they 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.

[34] Considering that the research is implemented by an Asian female researcher, it is not surprising to find that more minority founders are willing to participate in this research study.

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 are likely to be more pro-social and willing to help academic research studies. Second, our sample investors are likely to 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 are more likely to be interested in understanding how data-driven methods can help investment evaluations. Many investors also choose not to participate because they do not believe that an algorithm can help with the startup portfolio selection process if it does not quantify the founder's personality and the chemistry during an actual meeting.[35] Such sample selection bias does not hurt the experiment's internal validity, yet it implies that it is important to implement more experiments in different settings in order to check the external validity.

3.1.2 *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 Qualtrics survey online using their browsers. The survey tool contains the following two sections. After reading the consent form, investors will 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. In the second donation section (i.e., the dictator experiment section), investors will decide how much of an unexpected $15 Amazon Gift Card they want to donate to randomly displayed startup teams. Figure 3 provides the experiment flowchart that demonstrates the tool's structure.

A. Consent Form and Instruction Page

Both consent forms and recruitment emails invite investors to "try a matching tool that helps identify matched startups" and also notes that the anonymized data from investor responses will be used for studying investors' startup selection criteria, which is framed as secondary. Before the first profile evaluation section starts, I also provide an instruction page emphasizing that "the more accurately they reveal their preferences, the better outcomes the matching algorithm will generate (and the more financial returns that the lottery winner will obtain)" so that participants understand how the incentive works. Moreover, since most VC investors only invest in startups in their industries and stages of interest (called "the quality/disqualify test,[36] I ask all the participants to assume that the generated startups

[35] In this paper, I do not study the communication stage. However, Kanze, Huang, Conley and Higgins (2018) and Hu and Ma (2020) provide some insights on investors' behaviors in the communication stage.

[36] The first step of the investment process is to implement the "quality/disqualify" test before investors go through startup team composition and financial performance. The test, as a quick decision-making exercise, is based on many things such as the industry, stage, prior market knowledge, and other factors, which tell investors whether the startup is worth looking at. For example, an investor who invests exclusively in the B2B SaaS sector does not want to evaluate a healthcare startup. It is important to consider how to pass the "quality/disqualify" test when designing an IRR experiment as documented in Kessler et al. (2019) when they fail to replicate the IRR experiment at the University of Pittsburg.

they will be evaluating are in their industries and stages of interest.[37]

B. Section 1 (Incentivized Resume Rating Experiment)

B.1 Profile Creation and Variation

Following the factorial experimental design, multiple startup characteristics are dynamically varied simultaneously and independently, enabling me to test investor preferences of multiple important startup characteristics suggested by the existing theories.[38] I first create a set of team characteristics (including founding team's gender, race, age, education, previous experience, etc.), project characteristics (including market traction, comparative advantages, location, ESG criteria, etc.) 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.[39]

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).[40] In the same startup team, all the members are assigned names of the same gender and race to make such information more salient. Also, I emphasize the gender and race in both Q1 and

[37]Another potential way to pass the "quality/disqualify" test is to provide several survey questions asking the interested industry, stages, and even revenue range before the evaluation section as Kessler et al. (2019) did. For each different industry, researchers need to create different customized generated startup profiles which can capture the special characteristics in that industry by providing more details. I did not do this due to the following two reasons. First, the market changes very quickly in the entrepreneurial community. However, it usually takes a long time to prepare field work which needs the approval of IRB. Therefore, 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 happens often 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.

Given the restrictions mentioned above, I only choose to provide the information that is usually publicly available on LinkedIn, Crunchbase, or AngelList. Also, it provides the description of each hypothetical startup's comparative advantages. Some investors like Plug and Play Tech Centers sometimes go to these public platforms and look for relevant startups that fit their portfolios. The current design is to mimic this type of startup seeking behavior and provides data-driven methods for pre-selection decisions rather than fully substitutes the mainstream person-to-person deal flow process. Future researchers can think about more dedicated ways to pass the "quality/disqualify" test

[38]Introducing a rich set of randomly generated startup characteristics is usually not feasible in the correspondence test because of the following two reasons. First, an unusual combination of characteristics might raise investors' suspicions. Second, all the varied information inserted in the pitch email may not be salient enough. For example, it is a reasonable idea to randomize the traction or comparative advantages of a startup in the correspondence test. However, investors do not respond to such randomized information either because they feel this information is not verified and quite noisy, or because it is hard to compare the information with the benchmark because different founders may have different writing styles and some founders do not want to disclose too much information about their traction before they meet the investors. Therefore, as Bernstein, Korteweg and Laws (2017) mention in their paper, failing to find significant results related to project traction does not imply that the project does not play a role in the investment process.

[39]Sometimes the random combination may generate unusual cases like a startup with 50+ employees still not generating profits (see Amazon's history). Such cases account for a small percentage of total generated cases. However, future researchers can think about how to mitigate this issue when a rich set of characteristics are randomly varied and combined at the same time. It is helpful to collect as many of these uncommon cases as possible first to generate many filter criteria when writing the randomization code in order to capture the most common situations.

[40]Having a similar concern to Experiment B, I only added Asian entrepreneurs in the experiment because randomizing names is not suitable for testing other biases related to other ethnicity groups, like African American founders. In the U.S., African American founders and white founders have similar last name naming patterns, so I cannot use the last name to indicate race. African American founders and white founders have very different first name naming patterns, which makes it hard to use the first names to separate the effect of gender and the related social status and background.

Q2 by mentioning the founder's name again and using indicative words like "she/her/his/him/he." The list of full names used in the tool is provided in Table B1.[41] Similar to other components, the combination of first names and last names is dynamically implemented by Qualtrics.[42]

Manipulating Age and Education. — The age of the startup founder is indicated by the graduation year from their college or graduate school rather than being listed directly.[43] 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,[44] so the approximated age is calculated by the formula: age = 2020−graduation year+23. The randomization details are provided in Table 3. I also randomize the educational background between attending prestigious universities and more common universities, and a list of these schools is provided in Appendix B Table B2. All the universities selected have alumni who are real, successful startup founders based on the biography information recorded in the Pitchbook Database.

In order to generate relatively reasonable startup profiles, I implement the following three designs. First, each hypothetical startup profile is constructed using different components with ranges based on data from Pitchbook Database. Second, the information provided follows a format similar to Crunchbase's format and captures most of the online, publicly available information of each startup.[45] I did not provide extra, more private information like equity sharing plans because such information is generally not disclosed to the public in the pre-selection stage and is usually determined after several rounds of negotiation between investors and startup founders. Third, I also introduce a short break after investors evaluate the first half of the startup profiles (i.e., the first 8 profiles) by providing them with a progress screen and a startup ID for each profile to indicate the evaluation progress. This break is usually designed for testing implicit bias. All the randomization of different startup components, including startup team and project characteristics, is provided in Table 3. The detailed startup characteristics construction process is provided in Appendix B.

B.2 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 enough, I usually ask a

[41]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.

[42]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.

[43]It is suspicious to list age directly in a startup profile because none of the public startup platforms so this. Considering that age discrimination is a sensitive preference question, I use the graduation year as a proxy for age at the cost of accuracy in order to achieve more realism.

[44]Using 22 gives similar results. The reason why I use 23 is considering that some investors may assume the founders graduate from graduate school rather than college of these universities.

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

probability or percentile ranking question rather than a Likert scale question,[46] which has two advantages. First, these questions are more objective than Likert Scale questions. Second, the wide range from 1 to 100 provides richer and more detailed evaluation results and additional statistical power. This question design 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 Appendix B Figure B4 and Figure B5.

Mechanism Questions

The three mechanism questions are designed to test the following three standard, belief-based sub-mechanisms, which can potentially explain why investors care about certain startup characteristics. First, some startup characteristics can be indicators of the startup's future financial returns. To test this mechanism, investors need to evaluate the percentile rank of each startup profile compared to their previously invested startups, which is the quality evaluation question ($Q1$). Second, some startup characteristics may be suggestive of the startups' willingness to collaborate with certain investors rather than using other financial tools for their fundraising purposes, which is the "loyalty" evaluation question ($Q2$). Similar to the marriage market, the entrepreneurial financing process is also a two-sided matching process. Therefore, this type of "loyalty" potentially also matters. To test this channel, investors need to evaluate the probability that the startup will accept their investment rather than other investors'. Third, investors may use certain startup characteristics as indicators 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,[47] which is the risk evaluation question (Q_5).

The risk evaluation question is added when I recruit investors using only the matching incentive for robustness test purposes. During the recruitment process, I received feedback to add this question from several investors. Therefore, when recruiting the rest of the investors using only 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 important mechanism.[48]

Q1. (Quality Evaluation, First Moment) Imagine that [Founder Name] 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?

[46] 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 in their lab-in-field experiment.

[47] 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 or not, which is discussed in detail in Neumark (2012) and Heckman (1998).

[48] Similar to evaluating variance when testing discrimination in the labor market, obtaining investors' evaluations of risks for different startups is difficult using traditional empirical methods. However, considering its importance, such a mechanism is important to test if researchers need to fully understand an investor's investment decisions. An alternative way to obtain such information is to implement a new field project (for example, send an extra survey) as done by Bartoš, Bauer, Chytilová and Matějka (2016). However, since the alternative method cannot guarantee to collect information from the same group of investors, I decided to add such a question after adjusting the pre-registration plan and making modifications to the IRB proposal before implementing this change.

0 (extremely low quality) —— 100(extremely high quality)

Q2. (Collaboration Likelihood Evaluation, Strategic Channel) Considering the potential network and negotiation power of [Founder Name] 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.)?

0 (guaranteed rejection) —— 100(guaranteed acceptance)

Q5. (Risk Evaluation, Second Moment) 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)?

0 (No risk) —— 100(Highest risk)

Decision Questions

The two decision questions are designed to examine how the investors' preferences evolve from initial contact interest to investment interest. A standard experimental methods, like the correspondence test, generally observes the initial contact interest from candidate evaluators. However, it is still unknown whether contact interest can fully transform into investment interest or affect any later stage decisions. Therefore, I ask each experiment participant to indicate both their contact interest ($Q3$) and 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.

Q3. (Contact Interest) 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?

0 (will not ask) —— 100(will ask)

Q4. (Investment Interest) 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 in the team, drag the bar to 0.5.)

0 —— 1.0 (benchmark) —— 2.0 (>2.0)

C. Section 2 (Background Questions)

At the end of the matching tool, I also collect standard background information about the participant to check how representative my sample investors are and implement potential heterogeneous effects based on predetermined investors' characteristics. Such background information includes investors' preferred industries, stages, special investment philosophies (for example, only investing in social ventures and women-led startups) and standard demographic information, which includes gender, race and educational background. 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.

D. Section 3 (Donation Section - Dictator Experiment)

In order to directly identify taste-based mechanisms, I inserted a donation game at the end of the survey tool. Before they finish the survey, each investor will be informed that they will receive an unexpected \$15 Amazon Gift Card to thank them for participating in this research project.[49] However, they can also decide whether they want to donate a proportion of the provided \$15 to certain types of startup teams (i.e., if they donate \$3, they will receive a \$12 Amazon Gift Card). I will use the donated money to purchase a small gift for the corresponding type of startup founders in our collaborative incubators to bring them anonymous support and encouragement.[50] Each investor will see the following donation question:

"Thank you for completing the questionnaire. We will provide you with a \$15 Amazon Gift Card within 2 days. However, you can also choose to donate a portion of this \$15 to our Women's Startup Club to show your encouragement and 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 female startup founders.) Please choose how much you want to donate.
(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 female startup founder in our incubators to give them your anonymous encouragement.)"

The characteristics (i.e., gender and race) of the startup founders receiving the small gift are randomized, and both the pictures displayed and the wordings used in the description are changed accordingly. The options investors may randomly be provided with include 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 more salient, I also add a picture containing four startup founders of the same gender and race to make sure that survey participants understand what type of founders they are donating to.[51] All individuals in the pictures were smiling and professionally dressed to make sure

[49] I do not want to pollute the incentive structure designed for this experiment. Therefore, the compensation with the \$15 Amazon Gift Card is mentioned only at the very end of the survey tool and is not mentioned in either the consent form or the recruitment email.

[50] 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.

[51] The concern of using pictures in the experiment is that the appearance or other messages delivered by the pictures cannot be fully controlled. To mitigate this issue, I use four founders' picture 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 copy right problems. The information delivered

they are as much on equal footing as possible. Example founder's picture is provided in Figure 4.

3.1.3 *Incentives*

As an incentivized preference elicitation technique, the key point of the IRR experimental design is notable in that when subjects evaluate randomly generated hypothetical startup profiles, they understand that the more accurately they reveal their preferences, the more benefits they will obtain based on the incentive provided. Therefore, for all investors, I provide a "matching incentive" used Kessler et al. (2019). To increase the sample size, for a randomly selected subset of investors, I provide both the same "matching incentive" and a "monetary incentive" as used by Armona et al. (2019). Considering the amount of time required for participating in this experiment,[52] most participants should value the incentive. The details and justifications of both incentives are provided in the following two subsections.

A. Matching Incentive

For the randomly selected 4,000 investors who receive the recruitment email (Version 1), I only provide a "matching incentive," which means that after each investor evaluates 16 hypothetical startup profiles, we use a machine learning algorithm to identify matching startups from our collaborative incubators who will contact them for a potential collaboration opportunity if they are also interested in the investor's investment philosophy. The matching algorithm uses all of their evaluation answers to identify their preferences for different startup characteristics similar to Kessler et al. (2019). Therefore, all five evaluation questions are incentivized by providing this incentive, and a description of the algorithm is provided in the consent form.

The matching incentive has the following three merits. First, it can be applied to any two-sided matching market, such as the entrepreneurial financing market and the marriage market. Second, it can be used to 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, such an incentive often requires researchers to have certain social resources and connections in order to implement it.

B. Matching Incentive + Monetary Incentive

In order to increase the sample size, I provided both a "matching incentive" and a "monetary incentive" to a randomly selected 14,000 investors who received the recruitment email (Version B). Following Armona et al. (2019), the "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. Based on this monetary

by the pictures is more salient than that delivered by words.

[52] Some may be concerned that there are potentially two alternative motivations for investors to participate in this experiment. The first alternative incentive is to understand the algorithm and research method I am using for this matching tool. For such investors, the optimal decision is to read the consent form, evaluate a few startups, and then stop because the evaluation process is repetitive and time-consuming. The second alternative incentive is that some investors are very pro-social and willing to help research on entrepreneurial activities. However, this survey tool takes at least 20-30 minutes to finish, and some investors even replied to me that they would love to participate only if they are provided \$5000 as a consulting fee. Therefore, none of these alternative motivations should be a serious concern.

incentive, the more accurate their evaluations of each startup's quality are, the bigger the financial return they will obtain as a lottery winner.[53] 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 both the two investors who were randomly chosen to receive the award at the end of July 2020. The evaluation algorithm is provided in the consent form (Version 2).

Such a monetary incentive has the following merits and limitations. First, it mimics the real investment process in which investors have a certain amount of principal, and need to evaluate different startups accurately to generate maximum return. Second, compared to the matching incentive, such a monetary incentive does not require too many social resources and is easy to implement. Third, this incentive can be applied to more general situations besides a two-sided matching market. However, the current version cannot incentivize all the evaluation questions. When designing this monetary incentive, only the evaluation of startup's quality (i.e., Q1) is incentivized to avoid distorting participants' evaluations on other questions.[54] The related incentive structure for each evaluation question is provided in Table 4. Both incentives impose costs for making inefficient and inaccurate evaluations.

C. Justification

One concern of adding the "monetary incentive" to increase the sample size is that it will attract participants who do not value the matching incentive, which can generate noisy outcomes for Q2, Q3, and Q4. To justify the validity of adding the "monetary incentive," I have compared the evaluation results of investors who receive only the "matching incentive" and those who receive both the "matching incentive" and the "monetary incentive." The comparison results are provided in Appendix B Table B5, which shows that such a concern does not seem to be a serious issue because the interaction terms between the incentive structure and startup's gender and race are not significant.[55] Considering that one more question is added for Version 2 (with only the matching incentive), I did not compare the time spent on profile evaluation from investors recruited using these two incentive structures. Moreover, this experiment discovers multiple highly significant startup team and project characteristics that are crucial for investors' investment interest (see Table B4), which shows that investors understand the incentive structure and evaluate all the questions carefully.[56]

[53]For example, Peter Smith participated in this survey study and was chosen as one of the two lucky draw winners. In his survey, he indicated that on average, he felt that male teams are of higher quality and more likely to generate higher financial returns. In that case, we would then construct a portfolio containing more real startups with male teams. After one year, based on the financial performance of the portfolio on the Pitchbook Platform, this portfolio containing more startups with male teams generates a 10% return. So Peter Smith would receive \$500 + \$500*10% = \$550 as his finalized monetary compensation one year after he participated in the survey. \$500*10%=\$50 is the "extra monetary return." The historical return of the VC industry is between -15% and +15%, which means that the range of expected monetary compensation is roughly between \$425 and \$575.

[54]If the collaboration likelihood (i.e., Q2) is added to the financial return algorithm, then all the participants will 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 would be motivated to distort their true evaluation in order to maximize their financial return because both contact interest (i.e., Q3) and investment interest (i.e., Q4) can be affected by the collaboration likelihood (i.e., Q2).

[55]Although investors receiving the pure matching incentive are more friendly to older founders, this implies that the monetary incentive is potentially noisier than the pure matching incentive, and the bias detected in Experiment A is likely to be the lower bound of true bias.

[56]Another method to check whether participants understand the incentive structure is to separately ask them some questions that can test their understanding of this experiment. See Casaburi and Willis (2018).

3.2 Results

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

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

X_{ij} represents any founder's demographic information, like whether the founder is female or not. α_i are investor (i.e., subject) fixed effects that account for different average ratings across investors. Please remember that each type of startup characteristics is randomized orthogonally and independently. Therefore, the coefficient of this regression can be interpreted as causal evidence.

3.2.1 ***No Group-level Bias Against Female and Asian Founders.***

Table 5 reports regression results for group-level bias based on founder's gender and race. I use the total 1216 profile evaluation results, including both the first half and the second half of profiles. Panel A shows investors' attitudes based on the founder's genders. 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 the founder's race. Asian Founder is a dummy variable that is equal to one if the startup founder has an Asian last name, and zero otherwise. In Column (1), the dependent variable is the quality evaluation, which indicates the percentile rank of each startup profile compared to the startups previously invested in by the investor in terms of its potential financial returns. In Column (2), the dependent variable is the loyalty evaluation, which indicates how likely the investors think the startup team will accept their investment rather than the investment of 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 investment interest, which describes the relative investment amount compared to 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 that 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 to startups previously invested in by the investors in terms of its risk level. All the regressions include the investor fixed effect to control for any subjective judgement by each individual investor. Therefore, I compare profile evaluations within each individual investor. All the standard errors are robust standard errors.

The ordinary least squares (OLS) regression results show that all the coefficients on founder's gender and race variables are not significantly different from zero in both Panel A and Panel B, suggesting that there is no group-level bias against female and Asian founders. The null results potentially stem from the following three reasons. First, the results may suffer from the consent form effect, even though all the investors know that researchers will only use the anonymized data for research analysis, and they have incentives to reveal their true preferences. Investors understand that they are participating in a research project. Hence, their responses may potentially be more friendly towards minority founders. Second, the results may suffer from the sample selection bias problem in that recruited investors are more likely to be pro-social and willing to support the research studies implemented by female and Asian researchers. Third, compared to other other startup characteristics, which are more effective indicators of the expected financial returns and risks, founder's gender and race are no longer the first-order characteristics that profit-driven investors pay attention to. The first two reasons imply that the group-level bias found in this paper is likely to be the lower bound of the true bias against minority founders in the real world, and the preference towards female and Asian founders found in Gornall and Strebulaev (2020a) disappears in the experimental setting where the incentive to reveal true preferences is stronger.

3.2.2 *Belief-driven Implicit Bias Against Female and Asian Founders.*

Table 6 reports regression results that test whether investors have implicit bias against female and Asian founders by comparing their evaluation results in the first half of the study before the break with their evaluation results in the second half of the study after the break. Panel A tests the implicit bias based on the founder's gender. Panel B tests the implicit bias based on the founder's race. "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. 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, loyalty evaluation, contact interest, investment interest, and risk evaluation separately. All the regressions include the fixed effects for subjects, which allows some investors to be more generous than others with their evaluations.

Regression results of Table 6 show that when investors become fatigued or rushed in the second half of study, their evaluations of women-led and Asian-led startups significantly decline compared with men-led and white-led startups, which suggests the potential existence of implicit bias. Column (1) shows that investors spent 27 seconds fewer on evaluating each profile in the second half of the study after the break, while the average time spent on each profile is 59 seconds. This significant difference in response time implies that investors are fatigued or rushed after evaluating the first eight profiles, and hence they are more likely to use their unconscious judgement to evaluate each startup profile in the second half of study. In Panel A columns (2) and (5), the interaction effect of being a female founder and whether the profile is in the second half of study is significantly negative, indicating that investors' rating results of female founders' startups gradually decline compared with male founders' startups when they are fatigued. The p-values of "Female Founder" in the second half of study are about 0.11 and 0.12 in columns (2) and (5), indicating a weak group-level implicit bias. As Table 7 shows later, this gender bias is mainly driven by investors working in the tech sectors. Similarly, in Panel B column (2), the interaction effect of being an Asian founder and whether the profile is in the second half of the study is also significantly negative. Although the p-values of "Asian Founder" in the second half of study are about 0.2, Table 8 later shows that in the "high contact interest" situations, there is a strong implicit bias against Asians.

Table 7 tests implicit bias based on founders' gender for investors working in tech sectors and non-tech sectors. According to a large literature discussing gender issues in sciences or STEM industries (Carrell, Page and West (2010), Goldin (2014), and Kessler et al. (2019)), it is possible that investors from tech sectors have more implicit bias against female founders. To test this hypothesis, Penal A of Table 7 focuses on investors working in the tech sectors, which mainly include the information technology industry. Results of Panel A columns (2) and (5) show that investors, who are interested in the tech sectors, have stronger implicit bias against female founders based on their quality evaluation results and indicated investment interest. However, Results of Panel C show that investors interested in non-tech sectors, 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 bias against female founders from the tech-sector investors, Table 7 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 evaluation by 8.78 percentile rank. Then the implicit bias against female founders accounts for 44% of the Ivy League College effect. Column (4) shows that for the contact interest, the implicit bias against female founders accounts for 37% of the Ivy League College effect. Column (5) measures the investment interest, indicating that the magnitude of this implicit bias against female founders accounts for roughly 60% of the Ivy League College effect. This relative magnitude is not trivial, suggesting that the implicit bias could play an important role in tech sectors investors' decisions.

Table 8 tests investors' implicit bias based on the founder's race in the "high contact interest" situations and the "low contact interest" situations. Based on the distribution effect in Figure 9, which is discussed in detail in section 5, investors have implicit bias against Asian founders in the "high contact interest" situations (i.e., the likelihood of contacting the startup Q3 is greater than or equal to 50%). Results of Panel A in Table 8 show that the implicit racial bias against Asians mainly comes from the "high contact interest" situations. Columns (2), (4), and (5) show that in the second half of the study, Asian founders receive significantly lower ratings than white founders in the profitability evaluations (Q1), contact interest (Q3), and investment interest (Q4). However, results of Panel C show that in the "low contact interest" situations (i.e., the likelihood of contacting the startup Q3 is smaller than 50%), investors do not have this implicit bias.

Table 8 Panel B discusses the magnitude of this implicit bias against Asian founders in the "high contact interest" situations. Similar to Table 7 Panel B, I focus on the second half of the study, and use the effect of going to an Ivy League College as the benchmark. Panel B column (1) shows that the implicit bias against Asian founders accounts for 49% of the Ivy League College effect. Column (4) shows that for the contact interest, the implicit bias against Asian founders accounts for 38% of the Ivy League College effect. Specifically, column (5) finds that for the investment interest, this bias is roughly equal to 80% of the Ivy League College effect. Since the VC industry generally only invests in the top startups, the implicit bias against Asians in the "high contact interest" situations can potentially lower the probability of successful fundraising from the private equity market of Asian founders.

3.2.3 *Older Founders Are Considered to Impose Less Risk.*

Panel C of Table 5 reports regression results for the effect of founder's age on investors' profile evaluation results. Age is the approximated founder's age based on the graduation year from the college as indicated in section 3.1.2. Age^2 is the square of the founder's age. Fixed effects for subjects are included in all specifications. Columns (1)-(5) show the quality evaluation, loyalty 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 of Table 5 Panel C show how investors' evaluations respond to the founder's age. In columns (1)-(4), all the coefficients of "Age" and "Age^2" are not significantly different from zero. However, column (5) shows that the effect of age follows a convex pattern, indicating that, generally, older founders are considered to impose less risk. Results of Table 6 Panel D show that the belief that older founders are less risky is even stronger in the second half of the study. However, this experiment does not find group-level bias based on quality evaluation, contact interest, and investment interest even though older founder's startups are evaluated as involving less risk.

3.2.4 *Donation Results: Male Investors Are Less Likely to Provide Support to Female Founders.*

To test the potential taste-driven bias based on the startup founder's gender and race, Table 9 reports the regression results from the donation section (i.e., the dictator experiment) in which investors' donation behaviors does not help improve investors' social image or provide any contact opportunities. 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 is an indicative variable that equals to one if the displayed startup founder is female, and zero otherwise. "Asian Founder" is an indicative variable that 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 that are equal to one if the investor is female or Asian. All regressions use robust standard errors reported in parentheses.

Column (1) and column (4) show that investors hold a taste towards Asian founders, which makes sense because these experiment participants are likely to be more pro-social. Column (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 when all the support is anonymous. However, column (5) shows that female investors are significantly more likely to donate money to female founders even when all the support cannot be observed by founders and they will sacrifice the real monetary award. The results are consistent with the homophily channel that female investors are more likely to support female founders.

3.3 Discussion

Experiment A, which combines the IRR experimental design and the dictator game, has the following four merits. First, it provides a stronger incentive to reveal investors' preferences compared to the correspondence test method using the email setting. Second, it is extremely powerful for directly testing belief-driven mechanisms and taste-driven mechanisms. Also, the evaluation questions are carefully designed to provide insights on later stage decisions and allow the testing of decision-based heterogeneous effects (see Section 6) by generating rich evaluation outcomes within each individual. Third, the design is more ethical because it provides real benefits to experiment participants. Lastly, by allowing each individual to provide multiple evaluations, the experimental design generates a large enough sample size from a limited number of participants. In the entrepreneurial finance setting, senior VC investors are very hard to recruit, especially during the recession when most people focus on surviving the economic difficulties. Hence, the design is able to study experiment subjects that are generally harder to recruit by using other method.

However, when implementing such lab-in-field experiments, it is important to take note of the following limitations. First, similar to any experiments requiring voluntary participation, this lab-in-field experimental design also has potential sample selection bias during the recruitment process. It does not hurt the internal validity of the experiment. However, it is important to check the external validity by running a complementary experiment. Future researchers can also replicate this experiment in different settings by recruiting different investor groups. Moreover, any recruitment process which increases the response rate (i.e., collaborating with prestigious institutions, recruiting investors face-to-face, etc.) is helpful to mitigate such sample bias. Second, when testing sensitive preference questions (i.e., discrimination), the consent form potentially could affect participants' behaviors due to the observer effect even when all the evaluation questions are fully incentivized and only the de-identified data are used for research purposes. This implies that any detected bias is likely to be the lower bound of the existing bias in the real world. It is helpful to have future researchers test how strong the consent form effect is in such an incentivized experimental settings. Third, the incentive structure used in the experiment requires more social resources, which may not be user-friendly to junior researchers without many social connections. Therefore, any innovation on providing cheaper incentive structures is important to lower the experiment's cost. Lastly, except for the donation section (i.e., the dictator game), this experimental design does not generate real economic outcomes as a preference elicitation technique, which makes it hard to implement welfare analysis. Any attempts to obtain real economic outcomes or test later communication behaviors are important to help better understand the entrepreneurial financing process.

The lab-in-field experimental design in this paper can also be improved in the following ways based on investors' valuable feedback. First, the survey tool can be shorter if possible to recruit a larger number of investors. Second, it is helpful to ask for some simple investment criteria (i.e., revenue range, industry, etc.) before the evaluation section to make each profile more customized to different investors. For example, series B investors can evaluate more mid-stage companies. Such a design can improve the user's experience with the survey tool although it costs researchers more time and effort to implement. Third, more relevant information can be provided to investors, which includes founders' experience in the industry, whether previous startups succeeded or not, more background of existing investors, monthly burn rates, and more. Also, any efforts to improve the realism of each startup profile is also helpful. However, researchers need to be aware that too much information provided may dilute investors' attention, which makes it harder to test major variables of interest. Fourth, considering that most investment decisions are made on a relative basis at a specific moment, future researchers can ask each investor to compare multiple startups at the same time rather than to evaluate one by one. This "Netflix style" evaluation method is a more realistic way to capture such a relative investment strategy. A better format to visualize startup information and questions at the same time would also be helpful. Lastly, researchers can ask more questions about investor types, such as whether they are financial investors or strategic investors.

Before the academic community begins to widely use the recent new RCT method, as implemented in Experiment A, it is important to compare its results with the standard RCT method (i.e., the correspondence test), which is stronger in its external validity. Also, it is important to check whether the surprising results from Gornall and Strebulaev (2020a) are replicable and to further test the underlying discrimination sources in the cold call, pitch email setting. Therefore, I follow up with Experiment B, which uses the standard RCT method with a relatively advanced design.

4 Experiment B: Correspondence Test

In Experiment B, I study gender discrimination (male vs. female founders) and racial discrimination (White vs. Asian founders) in the VC industry of mainly the U.S. and other English-speaking areas in the world by sending randomly generated pitch emails and tracking detailed information acquisition behaviors of investors.[57] Specifically, to identify potential sources of bias, I introduce variation in startup quality (i.e., startup characteristics that affect investors' contact interest) in each email's subject line and the email's contents. This redesigned correspondence test follows a factorial experimental design, which orthogonally randomizes the startup founder's gender, race, educational background as well as the startup project's comparative advantages.

Sending out cold call pitch emails to investors for fund-raising purposes has become more popular recently following the trend of removing barriers to funding. For example, deck sender,[58] an online platform that helps entrepreneurs send pitch decks to the right investors for free, is designed to democratize access to funding and has sent out 90,000+ decks as of 06/2020. Compared with the mainstream fundraising methods from warm networks and in-person interactions, sending cold call emails does not require startup founders to establish close connections with practitioners in the VC

[57] These areas include UK, Canada, Australia, Singapore, Hong Kong, Isreal, India, etc. Considering the global trend in entrepreneurial activities and VC investment activities, I also recruit investors from other English-speaking areas. However, I do not include investors from China, Korea, and Japan due to the language used in this experiment, and I also do not recruit European Union investors because of the EU General Data Protection Regulation (GDPR).

[58] https://decksender.com/

industry,[59] hence lowering the entry cost. To some extent, this helps to increase the diversity of the entrepreneurial community.[60] However, considering the potential risk of idea theft and the lower response rate,[61][62] I would recommend that young startups try the mainstream fund-raising methods first before resorting to this probability game.

For ethnic discrimination, I only focus on testing potential bias against Asian groups considering the special setting of entrepreneurial finance and the timing of this experiment.[63] The Asian population constitutes the largest ethnic minority group in the U.S. entrepreneurial community. As documented by Gompers and Wang (2017b), the percentage of Asians grew from 10% to 18% among new venture capitalists and from 5% to 15% among entrepreneurs entering the market from 1990 to 2015. Asians also contributing substantially to U.S. innovation activities. However, the global anti-Asian sentiment due to COVID-19,[64] especially during the period when the term 'Chinese Virus" was used in the U.S., potentially lowers investors' expectations of Asian-led startups' profitability. Additionally, these sentiments increase the concern of involved risks, making Asian entrepreneurs' situations even more difficult.

Section 4 is organized as follows. Section 4.1 introduces the experimental design, including the email sending process and the email behavior tracking techniques. Section 4.2 describes the sample selection, including the summary statistics of both the real investors and fictitious startup founders. Section 4.3 discusses the analysis results of investors' information acquisition behaviors. Section 4.4 disentangles the underlying mechanisms, and Section 4.5 discusses the limitations of this correspondence test.

4.1 Experimental Design

Manipulating Identity of the Entrepreneur. —In order to indicate the gender and race of the fictitious startup founders, I have first generated a list of common, ethnically neutral first names which are highly indicative of gender (male or female) and also a list of common last names which are highly indicative of race (white or Asian).[65] Considering

[59] Gompers et al. (2020) show that unsolicited approaches by founders account for 12% of early-stage VCs' deals, and the majority of deals (62%) still come from professional networks and referrals.

[60] One concern of investing through personal networks is that minority founders may face more financing difficulties due to the lack of connections.

[61] Thanks to the advice from a managing director participating in the experiment, who informed us of the risk related to sending out cold call pitch emails. Some investors who seem to be interested in cold call pitch emails are likely to be just fishing around to get undeveloped but decent ideas worth stealing. Therefore, startup teams should be aware of these risks before sending out large scale cold emails. However, connecting with investors within your own network from alumni, events, or friends after careful due diligence sometimes works well. See the discussions on Quora: https://www.quora.com/How-do-I-pitch-a-startup-idea-by-email

[62] Gornall and Strebulaev (2020a) show that the cold email response rate of angel investors and VC investors was 6% in December 2018 (economics boom). In this experiment, which was implemented between 03/2020 - 09/2020 (economic recession), the cold email response rate is 1.5%. This phenomenon is consistent with Howell et al. (2020), which documents that early-stage investors are significantly more responsive to business cycles than later-stage investors.

[63] Similar to Gornall and Strebulaev (2020a), I did not assign African American names and other minority names in this experiment for the following two reasons. First, African American entrepreneurs are underrepresented in the entrepreneurial community and account for less than 1%. Therefore, commonly used African American surnames are less likely to accurately indicate the ethnicity in this setting. Second, the different first name naming patterns used by African Americans compared with the majority white group potentially will signal both economic background information as well as ethnic information, which makes it harder to identify the ethnic effect. Examining why other minority groups are underrepresented in the U.S. entrepreneurial community is an important question but outside the scope of this paper.

[64] The New York Times, "An Asian-American Author Talks About Racism in the Pandemic" June 24, 2020.

[65] 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 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.

the factorial experimental design used in this paper, I assigned four co-founders for each created fictitious startup team,[66] which include a white female co-founder, a white male co-founder, an Asian female co-founder, and an Asian male co-founder. 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 the selected names, and I have deleted any ambiguous names. The name lists (see Table C3) used in the experiment and the name generation process details are provided in Appendix C.

In order to introduce meaningful variation in startup quality, I randomized the educational background of each startup team and the project characteristics in both the subject line and the contents of each email. For the control group, I do not mention the founders' educational backgrounds at all. For the treatment group, the email's subject lines and contents indicate that the startup team's members come from prestigious universities in the U.S.[67] 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."[68] Therefore, this experimental design identifies whether startup founders' excellent educational backgrounds and startup projects' impressive advantages attract investors more compared to average startups that do not mention these factors when sending cold call pitch emails to investors.

Manipulating Access to Information. —The randomization of startups' characteristics (i.e., founders' gender, race, educational background, and project's advantage) 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,[69] 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 introduction 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 content, the investor can decide whether to reply or forward the email to other related investors who are potentially also interested in the same pitch email.

[66]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. There are many online platforms to help find co-founders, for example, FounderDating, StartHawk, etc. See https://press.farm/10-best-websites-to-find-a-co-founder-for-your-startup/

[67]Prestigious universities used in this experiment include Ivy League Colleges, MIT and Stanford. In the first-round 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. However, investors did not appreciate such a "mixed Ivy League educational background" as strongly as the "pure Ivy League educational background". Therefore, in the second-round experiment implemented between 08/2020 - 09/2020, I only used Ivy League Colleges, MIT, and Stanford as the educational background in the treatment group.

[68]MOM is abbreviated form for "month over month" growth in finance.

[69]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.

To make sure that the i.i.d assumption holds for the experiment,[70] the randomization 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,[71] 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,[72] 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,[73] 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.[74] Each investor received 3 to 5 pitch emails between 03/2020-09/2020.

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,[75] who are usually young startup founders or members of startup-related clubs at Columbia and other Ivy League colleges who are interested in this research project. We use Wix, a commercial website builder, to make the related startup websites which are in the under-construction stage. I do not create any LinkedIn accounts for these hypothetical startup founders because the LinkedIn Community does not allow creating suspicious accounts even for research purposes. However, the believability concern should not affect the email opening rate and the email reading time although it may affect the response rate and the contents of email replies. The pitch email example is provided in Figure 6, and the website example is provided in Figure 7.

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.[76] 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

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

[71] 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.

[72] 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.

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

[74] 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.

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

[76] 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.

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.[77] 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,[78] 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 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,[79] 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.

Email Behavior Measurements—I tracked the following email behavior measurements, including both the new behavior measurements used for this paper and the behavior measurements used in previous correspondence tests. These measurements include the email opening status and the corresponding time stamp, the email staying time measured in seconds, the sentiment of the email replies analyzed through LIWC,[80] the click rate of the inserted startup websites, and the response rate and whether the response is a positive response or a negative response.[81] 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 Appendix C Table C5. The flow chart of the first correspondence test is provided in Figure 8.

[77] 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.

[78] 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.

[79] 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.

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

[81] Positive response indicates either a direct invitation to a call or interest in the pitch deck.

Alternative Experimental Design without Deception—This experiment follows the mainstream practice of the correspondence test, which usually involves deception. There are two alternative experimental designs without deception that I implemented in 2018, but both of them failed for different reasons. For the first experimental design, I collaborated with real startup teams with both male and female co-founders for this experiment. This design can lead to legal risks that result from working with real businesses. Therefore, researchers must be extremely careful in designing all the consent documents and consent procedures. Also, there are not many startup teams with both female and male co-founders in most industries. Therefore, the recruitment process is extremely slow due to difficulties in finding these startups while maintaining a relatively good coverage of all industries. For the second experimental design, I organized a startup Pitch Night in 10/2018 in which I invited multiple VC investors to evaluate eight real startup teams pitching their ideas.[82] In the formal invitation email for this event, I introduced multiple exogenous variations in each real startup team's characteristics and tracked investors' email behaviors. This design failed because investors' response rates and the inserted startup website's click rates were extremely low. However, these failures provide crucial insights for the current design of the correspondence test.

4.2 Sample Selection and Data

This correspondence test experiment has been implemented multiple times in order to test the external validity. The first round was implemented between 03/2020 - 04/2020 during the outbreak of COVID-19 around the world. During this period, President Trump started using the expression "Chinese Virus" on 03/18/2020 and stopped using this expression around 03/23/2020 to protect the Asian American community.[83][84] Accidentally, this correspondence test captures this special "Chinese Virus" period. Considering the unusualness of this period, I implemented another round of the correspondence test in 2020 during the economy re-opening when people were calming down after the COVID-19 shock. Running the correspondence test multiple time periods helps test the external validity since the group-level discrimination atmosphere can be affected by different social events as shown in this paper. However, the current version of the draft only shows results from the first round.

Investors recruited for this experiment 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 provides the industry distribution of the created hypothetical startups. There are, in total, 67 startup ideas created with more than 200 names used in order to make sure that all experimental results are not driven by any special names or startup ideas. These ideas cover the majority of mainstream industries that venture capitalists are interested in, which include Information Technology, Healthcare, Consumers, Energy, etc.

[82] One startup successfully received investment from an event participant. Based on the feedback from practitioners, participating in these person-to-person pitch events are more likely for startups to build connections with investors and receive funding compared with sending cold call pitch emails.

[83] See https://theconversation.com/donald-trumps-chinese-virus-the-politics-of-naming-136796

[84] See Forbes News "Trump Abruptly Stops Calling Coronavirus 'Chinese Virus' At Daily Press Briefing". However, this expression was used again in 09/2020.

Note that with the correspondence test design, I cannot observe later stage decisions, such as investment interest for each startup. However, 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 (e.g. venture capital investment setting). Hence, if no further bias exists in the later-round communication stages, the amount of attention measured is indicative of investors' internal preference. Detailed discussion about this limitation and its impact is provided in section 4.5.

4.3 Results

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

Table 10 Panel A summarizes investors' major information acquisition behaviors in the first-round correspondence test. On average, the pitch email opening rate is 12.03% and each investor spent roughly 24 seconds on 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,[85] which indicates that early-stage investors are sensitive to business cycles as documented by Howell et al. (2020). Therefore, traditional measurements used in the correspondence test, like the email response rate, do not generate enough experimental power during the Covid-19 Pandemic. All of our experimental results rely on the new behavior measurements created in this paper.

Table 10 Panel B reports regression results of global investors' email opening behaviors for randomized pitch emails in Experiment 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."[86] "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 first-round correspondence test. 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. "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. All the regressions include start-up fixed effects to control for any idiosyncratic characteristics of each start-up pitch email, like the business models, etc. Hence, I am comparing investors' email opening rates within

[85] Gornall and Strebulaev (2020a) documents that the cold call pitch email response rate is roughly 6% in 2018 when the economy was in good shape.

[86] I use the period 03/18/2020-03/24/2020 to increase the sample size. Also, most U.S. news started to report this decision starting on 03/24/2020. See Trump Says He'll Stop Using the Term 'Chinese Virus'.

the same start-up's pitch email, and all the results are similar after including investor fixed effects. Following Bernstein et al. (2017), the standard errors are clustered at the investor level to account for the correlated opening decisions across different pitch emails received by the same investor.

Results of Table 10 Panel B show that investors' email opening behaviors respond to the startup founder's gender, race and educational background, but do not respond to the startup project's advantages. Column (1) shows that a pitch email sent by a female first name raises the opening rate by 1%, which is statistically different from zero. This difference is between pitch emails sent by female first names and pitch emails sent by male first names. 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 a pitch email sent by an Asian last name raises the opening rate by 0.7% after President Trump stops using the wording "Chinese Virus," which 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 in which only a pure Ivy League educational background being 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, Asian, and well-educated founders in the pitch email setting.

4.3.2 *The Direction of Bias Depends on Timing.*

Table 10 Panel C reports regression results of how startup characteristics affect investors' staying time on each pitch email, which approximates the attention spent on randomized pitch emails by each investor. The dependent variable is the time spent on each pitch email measured in seconds. In columns (1) and (2), I included unopened emails and replaced their email staying time with 0 seconds. Considering the potential truncation issue,[87] I also reported the sub-sample of opened emails in column (3). 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 there was a temporarily stronger bias against Asian founders during the COVID-19 outbreak in 03/2020, which started to fade after 04/2020. 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 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

[87] 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 level will bias our results towards zero. Therefore, it is important to compare magnitudes of race effect when only opened emails are included in the regression.

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, accounting for a 12.5% decrease in the staying time. Columns (2) and (3) suggests that there is a bias against Asian founders in 03/2020 and the direction of bias flipped after 04/2020. This implies that the direction of bias in Experiment B also depends on timing and can be affected by certain big social events, like the COVID-19 outbreak.

However, the temporary bias against Asian founders in 03/2020 should be interpreted as the lower bound of the bias in the real world due to the following two reasons. First, as shown in Section 3.3, investors are generally more friendly to Asian founders when stakes are low (i.e., the pitch email setting) and are implicitly biased against Asian founders when stakes are high (i.e., mainstream fundraising situations). If bias against Asian founders is even found in the pitch email setting, this implies that the bias against Asian founders in other situations can be even worse. Second, although emails with well-educated founding teams and better-quality projects have more contents emphasizing these advantages, the positive coefficients are not statistically significant in all the regressions. This means that the effect of the founder's race should be salient enough in order to generate significant results in this noisy experimental setting. Both reasons mentioned above suggest an even harsher fundraising environment for Asian founders during the COVID-19 outbreak in the real world.

4.4 Mechanisms (Testing Theories of Discrimination)

This subsection mainly explains why investors are biased towards female and Asian founders in the pitch email setting of Experiment B.[88] The mainstream discrimination theories in the correspondence test experiment setting can be classified into the following three types (Bohren, Imas and Rosenberg (2019b)): belief-based mechanisms (section 4.4.1), taste-based mechanisms (section 4.4.2), and amplifying mechanisms (section 4.4.3). In section 4.4.4, I also discuss other alternative mechanisms related to this specific experimental setting. Section 4.4.5 summarizes the detected 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 based on current theories. 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).

4.4.1 Belief-Based Mechanisms

Belief-based mechanisms (also called "statistical discrimination") are driven by differences in beliefs of a startup's productivity due to a lack of information (Phelps (1972), Arrow et al. (1973)). In the entrepreneurial financing setting, investors might prefer a certain group of startups based on gender and race potentially because such membership affects investors' beliefs about certain startups along the following three dimensions: the expected financial returns (first moment), the expected variance or risk of different groups (second moment) and the strategic channels (i.e., startups' willingness to collaborate in this two-sided matching financing process).

[88]For the mechanisms of why investors respond to founder's educational background, Bernstein et al. (2017) provide a detailed discussion. They find evidence that education is important due to the operational capabilities and expertise of the founders. The correspondence test setting does not provide direct evidence to separate the human capital effect (i.e., expertise) and the social capital effect (i.e., networks) from the founders' educational backgrounds.

Expected Quality and Financial Return (First Moment) Investors could prefer contacting female entrepreneurs and Asian entrepreneurs because they expect these types of startup founders to generate higher financial returns than the average. Investors may hold this belief because of previously documented facts,[89] the self-selection effect of minority founders,[90] different perceptions in the pitch email setting,[91] the lower negotiation power of minority founders,[92] more pleasant collaboration experiences,[93] etc. This paper focuses on testing whether investors hold the belief that minority founders can generate higher financial returns. I leave the work of investigating why investors have this belief and whether it is an accurate belief to future research.[94]

Table 11 shows that the racial bias towards Asian founders is mainly driven by this belief-based mechanism, while the gender bias towards female founders is likely not driven by belief. To differentiate belief-based mechanisms from taste-based mechanisms, it is important to have variation in startup characteristics that significantly affects investors' response (i.e., the educational background in this experiment). Since educational background only significantly affects investors' email opening rate rather than staying time, I use the email opening rate as the dependent variable to test this mechanism. Similar to Table 10, the dependent variable is equal to one when the investor opens the email, and zero otherwise. Column (4) shows that mentioning an Ivy League educational background reduces the bias towards Asian founders compared to white founders by 0.7% in opening rates, although the interaction effect of being an Asian founder and having an Ivy League educational background is not statistically significant. In order to increase the experiment's power, I focus on the sub-sample of emails which were sent after President Trump stopped using the phrase "Chinese Virus" on 03/23/2020 and emails whose treatment group mentions the "pure" Ivy League colleges in column (5). Results in column (5) show that sending pitch emails using an Asian name increases the opening rates by 2.6%, and mentioning an Ivy League educational background increases the opening rates by 3.0%. However, the interaction effect of using an Asian name and mentioning the Ivy League educational background is -3.2%, and all the coefficients are significantly different from zero. This means that after more information about the startup's team quality is revealed to investors, the bias towards Asian founders significantly shrinks. This is consistent with the belief-

[89] Ewens and Townsend (2020) show that female-led startups that investors connect with have higher chances of success than male-led startups on AngelList. Also, Fairlie and Robb (2010) show that U.S. Asian-owned businesses are more successful than White-owned businesses.

[90] Female entrepreneurs are more risk averse in general compared with male entrepreneurs. Therefore, for women who choose to start their own businesses, their startups could be of better quality and have better prospects in order for them to take this risk. Similar mechanisms also include cultural pressure (Fernandez and Fogli (2009)) and taste differences (Buttner and Moore (1997), Puri and Robinson (2013)). Also, minority founders, like women or Asians, need to overcome more difficulties in order for them to enter the high-impact entrepreneurial world due to a lack of connections or more challenges during school life. Baron, Markman and Hirsa (2001), Fryer Jr (2007), Bohren et al. (2019a), and Kacperczyk and Younkin (2019). Based on the dynamic discrimination theory (Bohren et al. (2019a)), the direction of discrimination can reverse if the investor's bias is based on incorrect beliefs. For example, minority founders may face more difficulties to get the same level of achievement due to discrimination against them. However, after they establish their quality, they are likely to be perceived as more capable than majority groups. See the female leadership premium documented by Rosette and Tost (2010).

[91] Investors may feel that unsolicited cold emails come from startups who have been rejected by the traditional investors within their network. Hence, these startups should have lower quality on average. Considering that minority founders have less access to traditional networks (Aldrich, Reese and Dubini (1989), Renzulli, Aldrich and Moody (2000), Shaw, Carter and Brierton (2001), Howell and Nanda (2019)), cold call pitch emails from minority founders send a smaller negative signal.

[92] Minority founders could have lower negotiation power if most investors are biased against them, which helps funds investing in them extract more surplus and benefits during the later round negotiation stage. Also, previous literature documents that women usually ask for less during negotiations. (Amatucci and Sohl (2004)).

[93] Shane, Dolmans, Jankowski, Reymen and Romme (2012) provides empirical evidence to show that Asians are easier to work with.

[94] Hu and Ma (2020), Bordalo, Coffman, Gennaioli and Shleifer (2016) and Bordalo, Coffman, Gennaioli and Shleifer (2019) have detailed discussions about the incorrect beliefs.

based mechanism because more signals about the startup's quality will correct the belief bias of each startup's quality that originates from a lack of information. If this bias is driven by a taste-based mechanisms, the interaction term should be insignificant or even positive because more signals about the startup quality cannot change investors' tastes if they just want to support the minority group. For example, in columns (1)-(3), I found that the interaction term of being a female founder and attending Ivy League colleges is insignificant and even slightly significantly positive. All the results above support that the bias towards Asian founders is mainly driven by the belief, while the bias towards female founders is likely driven by taste.[95]

Expected Variance of Different Group (Second Moment) Heckman and Siegelman [HS] (Siegelman and Heckman (1993); Heckman (1998)) sharply criticized the correspondence test by emphasizing that the expected variance of different groups can also affect the evaluator's decision. Based on the HS Critique, 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. This is because in the standard correspondence test setting, researchers only observe a nonlinear binary decision outcome (i.e., reply vs. no reply, opening emails vs. not opening emails, etc.) which can be affected by higher moment statistics. In a correspondence test design including only high-quality pitch emails, female founders can still receive more replies and attention from investors if they expect that 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. Neumark (2012) develops a model that can address this concern and recover an unbiased estimate of discrimination by including a meaningful variation in quality in the correspondence test. 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.

Table C6 shows that results are robust after correcting for the source of bias from unobserved variance using Neumark's model, which uses a Heteroscedastic Probit Model after imposing several parametric assumptions. Column (1) demonstrates that sending emails 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, which means that on average, investors expect that female founders to be more homogeneous. Column (2) and (3) show that sending emails 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 decreases starting in April. Although the expected variance of different groups is a potentially important confounding mechanism to test, it is not the main driver of the detected bias towards female and Asian founders.

Strategic Channel The entrepreneurial financing process in the VC industry is a two-sided matching process

[95] I do not make a firm, conclusive statement about the gender bias here because all the coefficients of using a female name and mentioning an Ivy League education are not significant here. The lack of strong significance can arise from the multicollinearity problem of the dummy variables in the regression. Hence, it is important to confirm this effect in the later rounds of the correspondence test. Also, the recent affirmative actions potentially make the team's racial information more informative of a team's quality than the team's gender information.

(Sørensen (2007)) where both investors and high-quality startups have bargaining power. Investors from small funds are likely to be more interested in minority founders if these founders may have stronger willingness to collaborate with them (i.e., loyalty) rather than other investors due to a lack of connections. In the standard correspondence test of the labor market, beliefs on the likelihood that candidates will accept job offers constitute a typical confounding mechanism, and "over-qualified" candidates can be rejected due to this strategic channel. It is an empirical question to test whether this strategic channel leads to the bias towards female and Asian founders.

Table 10 columns (4) and (5) show that the strategic channel is not the main driver of the bias detected in this correspondence test setting because mentioning an Ivy League educational background still significantly increases investors' email opening rates at the group level. If the strategic channel dominates, startup teams with excellent educational backgrounds should receive lower opening rates because they are "overqualified" and potentially less "loyal" due to more outside financing options. This is not surprising because investors usually expect startups sending cold call pitch emails are of relatively lower quality due to their lack of connections in the VC industry. Hence, startups are not likely to be "overqualified" in this specific experimental setting. The above results do not mean that "loyalty" does not matter because situations can change for high-quality startups who have much more bargaining power.

4.4.2 Taste-Based Mechanisms

Taste-based mechanisms include all the mechanisms that arises due to preferences and taste (Becker (2010)). In the entrepreneurial financing setting, investors have non-financial motivations for preferring a certain group regardless of the group's ability to generate higher financial returns. These taste-based mechanisms include friendly support of minority founders, homophily when investors and founders have a similar background, and miscellaneous mechanisms such as social image concern and others. However, the correspondence test experiment only provides indirect evidence to identify parts of the following subtle sub-mechanisms within the taste-based mechanisms.

Friendly Support Investors are likely to be biased towards minority founders because they want to support disadvantaged groups. For example, some VC funds or angel groups (e.g. 37 Angels) only invest in female-led startups. Also, with the growth of impact investment, more investors have begun to care about increasing diversity and meeting requirements from institutional limited partners with ESG goals.[96][97]

Table C7 shows that investors working in not-for-profit impact funds are more likely to be biased towards female founders and weakly biased towards Asian founders.[98] Column (1) documents that the interaction effect of using a female name and whether investors work in impact funds increases the email opening rate by 8.3%. Columns (2) and (3) show that using female names to send pitch emails increases the email opening rate by 10.3% for impact fund

[96] Global Wealth Advisors: https://www.gwadvisors.net/women-owned-businesses-esg-investing/ "Men rule the world, but not here" by Bloomberg News, January 24, 2020

[97] Environmental, Social, and Corporate Governance (ESG) refers to central factors in measuring the sustainability and societal impact of an investment in a company or business.

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

investors and only 1.1% for common funds which do not have special ESG missions. For both groups, the effect is statistically different from zero. However, the magnitude of this effect from impact funds is roughly 10 times that of the effect from common funds. This confirms the importance of impact funds in supporting female entrepreneurs, although they only account for a small proportion of all investors. Column (4) documents that impact funds are also more likely to open emails sent using an Asian name. However, the comparison of columns (5) and (6) show that the magnitude of this effect from impact funds is only 2 times the effect from common funds. However, all of coefficients are not significantly different from zero, which means most of the impact funds focus on improving gender equality. Table C8 shows that the heterogeneous effect between impact funds and common funds still exists for the "key word" classification method of impact funds.[99] Results above show that some parts of the bias towards female and Asian founders come from friendly support from impact funds.

Homophily Homophily means that people prefer the groups that share similar backgrounds as themselves. Based on the prediction of homophily theory, minority investors would prefer minority founders and majority investors would prefer majority founders.[100] Table 12 shows that this experiment finds a weak homophily effect based on investors' email opening rate and staying time. Interestingly, columns (2) and (3) show that sending pitch emails 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. However, 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. The coefficients are not significant in the noisy experimental setting, which does not mean that this homophily effect does not exist. 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. Results above suggest that male investors are more likely to open pitch emails sent by female names, but it is the female investors who spend more time on pitch emails sent by female names.

Miscellaneous Mechanisms Other taste-based mechanisms include the social image effect, which argues that providing help to minority founders potentially improves investors' social images. Another sensitive mechanism is the sexual harassment concern,[101] which has brought widespread attention to the treatment of women in the entrepreneurial community. Unfortunately, Experiment B does not provide direct evidence to identify these two mechanisms.

4.4.3 Amplifying Mechanisms

Mechanisms that can magnify both taste-based bias and belief-based bias also exist. These amplifying mechanisms include attention discrimination and implicit bias. "Attention discrimination" theory (Bartoš et al. (2016)) predicts

[99]Besides the funds whose primary investor type is not-for-profit, I also include funds whose stated investment preferences contain key words like "ESG," "impact," and "MWBE".

[100]Egan, Matvos and Seru (2017) document the patterns of "in-group" tolerance where managers are more forgiving of misconduct among members of their own gender/ethnic group in the financial advisory industry.

[101]"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.

that even if complete information about an individual is readily available, discrimination can still happen because investors may endogenously allocate their scarce attention to their preferred groups before they make their decisions. "Implicit bias" refers to the attitudes or stereotypes that affect investors' decisions in an unconscious manner.

Attention Discrimination Investors may pay more attention to their preferred groups before they make decisions in the entrepreneurial investment setting. Considering that all of our outcome variables (i.e., email opening rates, time spent on each pitch email) actually measure the attention of investors rather than the finalized decisions (i.e., investment decisions), results support the existence of the attention discrimination channel. If there is no additional bias or friction arising in the later communications stages, the amount of attention should be positively correlated with interest in later round decisions.

Implicit Bias Investors may make less careful judgements when they are fatigued. This means that even though most investors do not explicitly consider gender and race in their investment process, stereotypes can affect their judgement when they are busy or fatigued. Unfortunately, the correspondence test experiment does not provide direct evidence to test this channel because it is hard to observe investors' status when they open the pitch email even though I can observe the time stamp when each email is opened.

4.4.4 Alternative Mechanisms

Uninformative Email Replies Investors may reply more to minority founders and use a more positive and friendly tone in order to be politically correct. Therefore, such email replies are not indicative of their true preferences or imply investment decisions. However, this mechanism cannot explain the results found through measuring the email opening rates and email reading times because these behaviors are usually not measured in previous correspondence tests or observed by the founders directly. Therefore, this special mechanism, which is related to the pitch email experimental setting, can be ruled out confidently.

4.4.5 Summary: Taste-Based Gender Bias & Belief-Based Racial Bias

To sum up, Experiment B finds taste-driven bias towards female founders and belief-driven bias towards Asian founders. For the gender bias, this experiment documents the friendly support of impact funds and does not detect any belief-driven mechanisms. For the racial bias, this experiment documents that investors prefer Asian founders in the pitch email setting because they expect Asian-led startups to be of relatively higher quality. However, I do not find strong taste-driven mechanisms for the racial bias. It should be noted that attention discrimination exists in this experimental setting.

4.5 Discussion

Experiment B essentially sacrifices internal validity in exchange for stronger external validity. Therefore, it is important to realize the following limitations of Experiment B. First, the email experimental setting is very noisy because all email behavior measurements are not perfect and many factors (like weather and emotion) can affect how investors treat each cold call pitch email. This means that the correspondence test can only detect strong effects from randomized information and has trouble detecting other potentially existing effects. For example, results about randomized startup project characteristics are not significant. However, we cannot conclude that startup projects do not matter for financing outcomes here. The noisy setting also limits the number of research questions a researcher can test because many introduced variations are not salient enough to generate significant results. Future research can work on how to improve the measurements of investors' behaviors.

Secondly, the results from the pitch email setting may not be generalizable to other mainstream fundraising settings because the incentive to reveal investors' true preferences towards female and Asian founders is relatively weak in the email setting. Also, the experiment is implemented during the pandemic recession period. As documented in this paper, the discrimination atmosphere depends on timing and can be quickly influenced by big social events. Therefore, future research can study more mainstream entrepreneurial financing settings, like the warm introduction setting, and replicate this experiment during the economic boom when investors' response rates are likely to be much higher.

Thirdly, the correspondence test in Experiment B, as a preference elicitation technique, only observes initial contact interest rather than later stage investment interest. Also, I do not observe real economic outcomes. This limitation makes it hard to implement welfare analysis and transform various email behaviors into real money analysis. Not being able to observe later stages limits the mechanisms that I can directly test. For example, I cannot identify whether the social image effect or a sexual harassment concern exists in this experimental setting. Some research work (Hu and Ma (2020), Kanze et al. (2018)) analyze video data to study the communication stage. However, more quasi-experimental research in the future would be helpful to generate real economic outcome analysis.

Lastly, this correspondence test design involves deception similar to most standard correspondence test designs despite efforts to attempt two alternative ethical designs as discussed in Section 4.1. As pointed out by List and Rasul (2010), "ethical issues surrounding human experimentation is of utmost import."[102] After the data confidentiality rule was implemented in the European Union starting in 2018, an experimental method that does not provide a consent form to participants is infeasible in these areas. Hence, future research should continue the process of improving the correspondence test design and mitigating the deception issues involved.

Despite the limitations mentioned above, Experiment B has the following two merits. First, it has a relatively stronger external validity compared with Experiment A because it can recruit a much larger number of investors by sending randomized hypothetical pitch emails. Second, it also improves the internal validity compared with the standard

[102] For the practical guidance of field experiments, see Dunford (1990).

correspondence test design. Compared with the standard correspondence test design, it solves the "low-response-rate" problem by tracking investors' detailed email information acquisition behaviors and by introducing variation in the email's subject line. Therefore, it can test discrimination sources in the pitch email setting and relatively increases the internal validity compared with the standard correspondence test design.

5 Discussion: Reconciling Contradictory Results

5.1 Reconciling Contradictory Results

This section will reconcile the contradictory results of the previous two experiments and the previous literature by showing the effect of founders' gender and race across the distribution of an investor's contact interest.[103] Both Experiment A and the descriptive papers in the literature (Ewens and Townsend (2020), Henderson et al. (2015)) find that investors are biased against female and Asian founders. However, the correspondence tests of Experiment B and Gornall and Strebulaev (2020a) instead find bias towards female and Asian founders. All the regression specifications used before identify the average treatment effect of a founder's gender and race on investors' interest. However, as pointed out by Neumark (2012) and Heckman (1998), the magnitude and direction of these average preferences can vary across the distribution of the evaluator's contact interest. Therefore, to reconcile all the mixed results above, I take advantage of the detailed contact interest rating to demonstrate investor's bias across the distribution of contact interest.

Figure 9 demonstrates the effect of a startup founder's gender, race, and age across the contact interest distribution using the profiles evaluated in the second half of Experiment A. Given the potential consent form effect, profiles evaluated in the second half of this study are more likely to represent investors' true preferences.[104] Panel A provides the empirical cumulative density function (CDF) for a founder's gender on investors' contact interest rating (i.e., $Pr(\text{Contact Interest} > x|\text{Female Founder})$ and $Pr(\text{Contact Interest} > x|\text{Male Founder})$). Panel B provides the OLS coefficient estimates (i.e., $Pr(\text{Contact Interest} > x|\text{Female Founder}) - Pr(\text{Contact Interest} > x|\text{Male Founder})$) and the corresponding 95% confidence interval. Similarly, Panel C and E provide the empirical CDF for a founder's race and age. Panel D and F provide the OLS coefficient estimates for a founder's race and age.

Figure 9 shows that the direction and magnitude of bias based on a founder's gender, race, and age depends on context. When stakes are low, investors are biased towards female, Asian, and older founders.[105] However, when stakes become higher, investors are biased against female, Asian and older founders. In Panel A, when the contact interest rating, as measured by the probability of contacting the startup, is lower than 5%, the CDF for a female founder is slightly to the right compared with the CDF for a male founder. This means that in this situation, female founders are

[103] I use the contact interest rather than investment interest mainly because previous correspondence tests and Ewens and Townsend (2020) only observe the contact interest rather than investment interest. Therefore, using contact interest is more appropriate to reconcile the literature.

[104] Figure B1 in Appendix B uses the total profile evaluations and shows that patters are similar to that in Figure 9. However, the magnitude of bias is much smaller due to the potential consent form effect.

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

slightly preferred. In Panel B, the coefficient of being a female founder is also slightly positive on the left tail of the distribution. However, as investors' contact interest increases, male founders are gradually more and more preferred. In Panel A, it is clear that for most positions on the distribution, the CDF of a male founder is on the right side of the CDF of a female founder. In Panel B, the coefficients of being a female founder become gradually negative for most of the situations on the distribution. Figure 9 Panel C and D show a similar pattern that is even more salient for Asian founders. When the contact interest is below 25%, investors are biased towards Asian founders. However, when the contact interest is above 25%, investors are biased against Asian founders. Figure 9 Panel E and F demonstrate that this phenomenon also exists for older founders.

This "crossing" data pattern for gender and racial bias discovered in Figure 9 provides a crucial insight to reconcile the contradictory results of the previous literature and the previous two experiments in this paper. Correspondence tests using the cold call pitch email setting in Gornall and Strebulaev (2020a) and Experiment B mainly capture the left tail of the whole distribution. Compared with the mainstream fundraising method of networking or referral, sending cold call pitch emails is more likely to be used by startup teams without enough connections or strong background. Therefore, it is not surprising to see that this email setting mainly describes the situations in which investors do not have strong contact interest. However, previous descriptive papers and Experiment A mainly capture the middle and right part of the whole distribution where investors are more biased against female and Asian founders. Ewens and Townsend (2020) use the unique data on the AngelList platform, and Experiment A collaborates with real accelerators, which all focus on relatively more attractive startups. To sum up, previous literature and RCT experiments find different results mainly because they study investors' biases in different positions within the whole distribution. Unfortunately, discrimination results of gender and racial bias from correspondence tests cannot be generalized to the more mainstream fundraising setting in the entrepreneurial financing setting. If future researchers implement another RCT or quasi-experiment that studies high-stake situations, this paper predicts that they should be able to find bias against female and Asian founders.

Such "crossing" data pattern for gender bias can be explained by the following trade-off faced by investors. On the one hand, providing support to female founders can increase investors' internal self image or social image, which is proved by Experiment B. On the other hand, investors hold the belief that women's startups are less profitable (see Section 3.2 and Section 6), then preferring female founders potentially lowers their financial returns. This is proved by Experiment A. Therefore, when the cost of providing support is low in low-stake situations, investors can be more friendly and provide more advice or help. However, it does not mean that they would like to sacrifice their financial interests to provide this support in those high-stake situations.[106]

The reversion of racial bias can be explained by investors' belief patterns. For the relatively low quality startups, both Experiment A and B show that investors feel Asian founders' startups are more profitable in this situation. This is

[106] A simple structural model will be provided in the next draft. DellaVigna, List and Malmendier (2012), DellaVigna, List, Malmendier and Rao (2016) and Floyd and List (2016) provide good examples of how field experiments can provide exogenous variation to estimate structural models.

reasonable considering that Asian founders generally need to overcome more difficulties to achieve similar performance. Moreover, people have the stereotype that Asians are more hardworking. However, Experiment A shows that for the relatively high quality startups, investors generally believe that white founder's startups are more profitable. This is also reasonable considering that creating a unicorn startup requires top social resources and white founders can have more advantages in this situation.

The "crossing" data pattern for the bias based on founders' age is more complicated. Although Table 6 shows that older founders are considered to be less risky, risk can not fully explain this data pattern. Theoretically speaking, many other mechanisms can coexist, and Experiment A and B do not provide rich evidence to explain agism in this setting. I will leave it to future research agenda.

5.2 Experiment A and Experiment B Are Complementary

This paper shows that the recent RCT method (i.e., lab-in-field experiment used in Experiment A) and the standard RCT method (i.e., correspondence test used in Experiment B) are complementary rather than fully substitutive as suggested by Kessler et al. (2019) due to the following reasons. First, the lab-in-field experiment has stronger internal validity (i.e., it tests more mechanisms, and provides stronger incentive to reveal true preferences) and weaker external validity (i.e., recruits smaller number of subjects), while the correspondence test has stronger external validity (i.e., recruits a large number of subjects) and weaker internal validity (i.e., struggles to test mechanisms, provides weaker incentive to reveal preferences). Therefore, combining them together can provide a more complete picture of the discrimination phenomenon. Second, the lab-in-field experiment 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. Hence, comparing results from both experiments can also help researchers to better understand the magnitude of discrimination in different situations. Future researchers can make an effort to improve internal validity and external validity of both RCT methods. Also, it is worth creating alternative correspondence test experimental designs that do not involve deception.

5.3 Policy Implications

To mitigate the discrimination problems in the venture capital industry, the results of this paper's experiments provide the following three policy implications. First, increasing the diversity of the venture capital industry would be helpful. As documented in Experiment A, female investors are more likely to provide support to female founders than male founders. Moreover, Raina (2019) documents that syndicates with female lead General Partners are better able to evaluate female-led startups. Hence, increasing the diversity of the investment community can even be more profitable. Second, increasing impact funds aiming to support minority founders is crucial. As discovered in Experiment B, impact funds are much more likely to open female founders' pitch emails compared with profit-driven funds. Therefore, this group plays an important role in mitigating the discrimination issues in the investment community. Third, any training or actions that help mitigate the implicit bias of investors are important. This paper shows that belief-driven implicit bias is the main driver of the bias against minority founders, and implicit bias often occurs when investors are fatigued or tired. Thus, in the fundraising activities like Startup Pitch Night, it is helpful to provide minority founders with earlier time slots for their presentations when investors are not tired yet.

It should also be noted that discrimination problems can be more severe during a period of economic recession compared to an economic boom. Many venture capital funds use the relative investment strategy and select relatively more profitable startups during a certain period. During a recession, the VC industry is generally more selective and increases their bar for investment purposes. However, bias against minority founders can be larger in these market conditions. Therefore, it is important for the entrepreneurial community and policy makers to take actions to prevent more serious discrimination problems when the economy goes down.

Also, if the bias against minority groups exists in the whole economic system, from the early stage financing process to the later stage financing process, it will impose much more difficulties to correct such bias in the initial contact stage. In a systematically biased economic system, all the participants will be trapped in a stable equilibrium. For example, even some investors have no bias against minority groups, it is rational for them to not invest in such startups if it makes them harder to exit or sell the startup to later-stage biased investors. Therefore, it is important for future researchers to test whether other parts of the entrepreneurial financing system also suffer from discrimination issues. It is also crucial for the entrepreneurial community to unite together and provide extra supports and motivations to correct such biases.

6 Are We Still United, and What Separates Us?

The previous two experiments have discovered the following two opposing groups that both exist in the investment community: the pro-minority group investors (i.e. impact funds) and the anti-minority group investors. To test how divided the investment community is in terms of investors' attitudes towards minority founders and what separate us, I have developed a consistent, decision-based heterogeneous effect estimator.

This estimator can test what are the separate driving forces of the anti-minority groups and the pro-minority groups, which are defined by investors' indicated decisions rather than their pre-determined demographic information. In an increasingly divided society, groups potentially make opposing decisions based on different motivations. For example, pro-minority investors may prefer investing in minority founders for taste-based reasons. They may simply want to provide their support to women rather than to generate higher financial returns. On the other hand, anti-minority investors may prefer not to invest in minority founders for belief-based reasons, such as not believing that such founders' startups can be profitable. However, the group-level average treatment effect used in previous literature can only detect the dominant mechanisms rather than the separate driving forces of people making different decisions. Therefore, I have developed a decision-based heterogeneous effect estimator to fill in this gap.

The logic behind how this estimator works is very simple. Since Experiment A introduces within-individual level randomization and requires investors to evaluate multiple randomized startup profiles, theoretically speaking, researchers can identify whether each individual investor is a pro-minority investor based on his/her indicated decisions (i.e. contact interest and investment interest). Therefore, in an ideal situation, it is feasible to classify recruited investors into groups who prefer contacting (or investing in) minority founders and groups who prefer contacting (or investing in) majority founders. Researchers can then run separate pooled regressions within each group to investigate each group's mindset. However, to solve the potential generated regressor problems in a nonideal situation, I need to use the leave-one-out technique to create a consistent estimator. Detailed proof and discussion of this estimator are provided in Appendix E. It should be noted that the current version of this estimator still replies on the assumption of linearity and the more generalized form of this estimator will be provided in another paper.

Table 13 provides the decision-based heterogeneous effect for founders' gender, which measures the evaluation results of pro-women investors and anti-women investors who are defined by their indicated contact interests. Panel B describes the evaluation results of investors who prefer contacting female founders, and Panel A describes the evaluation results of investors who prefer contacting male founders. All the coefficients and standard errors in the parentheses are calculated using the "leave-one-out" estimator and the bootstrap method due to relatively small sample size. Similarly, Table 14 and Table 15 provide the decision-based heterogeneous effect for the founders' race and age.

I have found that female founders face a larger division in investor attitudes than Asian and older founders. The reason for this appears to be differing expectations of startup profitability. Table 13 column (1) shows that "anti-women" groups have lower contact interest and investment interest for female founders mainly due to belief-based reasons. These founders feel that women-led startups are less profitable, and have 16.40 percentile ranks lower potential financial returns 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. These "pro-women" investors believe the opposite and feel that women-led startups are more profitable. Therefore, the divisions in profitability expectations for women-led startups is one explanation for investors' different decisions on whether to contact female founders.

Table 14 and Table 15 show that for "anti-Asian" and "anti-older" groups, lower expectation of these founders' profitability is also an 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 on whether to contact these founders. However, for older founders, imposing less risk is another likely reason why "pro-older" investors prefer older founders, as shown in Section 3.2. For "pro-Asian" groups, taste is another potential reason because the donation results show that recruited investors in our sample are more friendly to Asian founders, which is driven by taste.

Altogether, the results show that the key for minority founders to improve their chances of obtaining investment from venture capital industry is to improve investors' expectations of their startup's future financial returns. This is especially important for attracting those "anti-minority" investors, whose decisions are mainly driven by their internal beliefs of startups' profitability. It should be noted that such beliefs can be right or wrong in the real world. Therefore, any type of training that helps minority founders improve their persuasion skills is potentially helpful for successfully raising funding.

7 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, there is scarce causal evidence to answer it due to data limitations on unobservable start-up characteristics and the lack of exogeneous variations to solve the endogeneity problems, not to mention test the underlying mechanisms. Moreover, previous literature provides the following contradictory results. The standard RCT method (i.e. correspondence test) proves there is bias towards minority founders, while descriptive papers demonstrate that there is bias against minority founders.

To reconcile the disparate results in the literature and solve the limitations of the standard RCT method, this paper implements the following two randomized controlled trials by recruiting real venture capitalists mainly from the U.S. Experiment A makes use of a newly created "Nano-Search Financing Tool" and invites U.S. investors to evaluate 16 startup profiles, which they know to by hypothetical, in order to be matched with appropriate startups from the collaborative incubators. Investors can also use the tool to donate a small amount of money to randomly displayed startup teams. Experiment B uses new email behavior tracking technologies and an advanced design to compare investors' reactions to hypothetical pitch emails with randomized startups' information.

Results show that the direction of bias depends on both context and timing. In relatively "low-stake" situations, investors are biased towards female, Asian and older founders. However, in relatively "high-stake" situations, investors are biased against female, Asian and older founders. Moreover, this paper finds a temporary, stronger bias against Asian founders during the COVID-19 outbreak, which started to fade after April 2020. Compared to female investors, male investors are less likely to provide anonymous support to female founders. Such bias stems from multiple sources, including implicit bias, attention discrimination, belief-driven mechanisms and taste-driven mechanisms. Specifically, statistical discrimination is an important reason for "anti-minority" investors' contact and investment decisions. These results reconcile the contradictory results in the literature by demonstrating how the biases vary in different settings and suggesting that different studies may only capture a specific part of a larger picture.

Overall, this paper confirms the existence of early-stage investors' bias based on startup founders' gender, race, and age using more advanced RCT methods. Hence, it contributes to the debate about discrimination in the venture capital industry and also the development of more powerful experimental tools in the experimental economics literature. It is important for future researchers to test whether such bias also exists in other parts of the entrepreneurial financing system and to implement welfare analysis using real economic outcomes.

References

Aldrich, Howard, Pat Ray Reese, and Paola Dubini, "Women on the verge of a breakthrough: Networking among entrepreneurs in the United States and Italy," *Entrepreneurship & Regional Development*, 1989, *1* (4), 339–356.

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.

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, *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.

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 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.

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.

Cen, Xiao, "Household Wealth and Career Choices: Evidence from Natural Disasters," *Working Paper*, 2020.

DellaVigna, Stefano, John A List, and Ulrike Malmendier, "Testing for altruism and social pressure in charitable giving," *The quarterly journal of economics*, 2012, *127* (1), 1–56.

——, ——, ——, and Gautam Rao, "Voting to tell others," *The Review of Economic Studies*, 2016, *84* (1), 143–181.

Dunford, Franklyn W, "Random assignment: Practical considerations from field experiments," *Evaluation and Program Planning*, 1990, *13* (2), 125–132.

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.

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," Technical Report, National Bureau of Economic Research 2017.

—— and ——, "Diversity in Innovation," Working Paper 23082, National Bureau of Economic Research January 2017. Series: Working Paper Series.

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

Gompers, Paul A., Will Gornall, Steven N. Kaplan, and Ilya A. Strebulaev, "How do venture capitalists make decisions?," *Journal of Financial Economics*, January 2020, *135* (1), 169–190.

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," SSRN Scholarly Paper ID 3318245, Social Science Research Network, Rochester, NY March 2020.

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

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.

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

List, John A and Imran Rasul, "Field Experiments in Labor Economics.‖ National Bureau of Economic Research (Cambridge, MA) Working Paper No. 16062," 2010.

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.

Renzulli, Linda A, Howard Aldrich, and James Moody, "Family matters: Gender, networks, and entrepreneurial outcomes," *Social forces*, 2000, *79* (2), 523–546.

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.

Shaw, Eleanor, Sara L Carter, and Jackie Brierton, "Unequal entrepreneurs: Why female enterprise is an uphill business," 2001.

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.

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 the active venture capitalists (defined as those whose email addresses are verified by the testing email) who received the cold call pitch emails in the correspondence test and the recruitment emails in the lab-in-field experiment. Panel A reports the geographical distribution of the sample investors. "Others" includes South Africa, Cayman Islands, Malaysia, and etc. Panel B reports the industries that these investors have stated that they are interested in investing. An investor can indicate multiple preferred industries. "Others" includes special industries like packaging technologies. 3.8% of the investors' industry preferences cannot be found online and I have assumed that they are interested in all of the industries when sending out pitch emails. Panel C reports the investors' demographic information and investment philosophy. '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 angel group, and zero otherwise. If an investor is both an 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 Pitchbook. 'Top University=1' and 'Graduate School =1' are indicator variables that equal one if the investor attended a top university (i.e. Ivy League Colleges, MIT, Duke, Caltech, Amherst, Northwestern, Stanford, UC Berkeley, University of Chicago and Williams College) or 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 Stated Investment Philosophy

	N	Mean	S.D
Cold Email Acceptance	69	0.74	0.44
Prefer ESG	69	0.17	0.03
Direct Investment	69	0.94	0.24

Continued

Panel E: Available Fund's Financial Performance

	N	Mean	S.D	Percentile 10	50	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 participated in the lab-in-field experiment. 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 the 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 their position is as an analyst (intern) or associate. "Cold Email Acceptance" is an indicator variable which equals to 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 to 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 Pitchbook.

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 collaborative 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 the collaborative 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 the 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 saw 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 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 5: Experiment A Evaluation Results About Gender, Race and Age

Dependent Variable	Q1 Quality (1)	Q2 Loyalty (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 tables describes evaluation results combining the total profile evaluations, including the eight profiles in the first half and the eight profiles in the second half. Some investors skipped the evaluation questions of loyalty or investment if they feel the information is not enough to make their judgements. Q5 (risk evaluation) is only added to a randomly selected investors for robustness check. Panel A shows investors' attitudes based on founders' gender. "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. "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 loyalty 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. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 6: Experiment A Implicit Bias Based on Founder's Gender, Race and Age

Dependent Variable	Response Time (Unit: Second) (1)	Q1 Quality (2)	Q2 Loyalty (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 Loyalty (3)	Q3 Contact (4)	Q4 Investment (5)	Q5 Risk (6)
Panel D: 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 × 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 × 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, loyalty 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. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 7: Experiment A Implicit Bias Based on Founder's Gender by Investors' Industry

Dependent Variable	Response Time (Unit: Second) (1)	Q1 Quality (2)	Q2 Loyalty (3)	Q3 Contact (4)	Q4 Investment (5)	Q5 Risk (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 graduates 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 95th percentile (59.23 seconds on average). Columns (2)-(6) show the quality evaluation, loyalty 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. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table 8: Experiment A Implicit Bias Based on Founder's Race by Investors' Contact Interest

Dependent Variable	Response Time (Unit: Second) (1)	Q1 Quality (2)	Q2 Loyalty (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, loyalty evaluation, contact interest, investment interest and risk evaluation separately. Standard errors in parentheses are robust standard errors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

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 the regression results from the donation section (i.e., the dictator experiment), 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 first-round 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 first-round 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 the interaction effects between the educational background of a startup founder's gender and race using investors' email opening rate as the outcome variable. 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 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 the underlying dominant mechanism, 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 first-round 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 combined 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 the startup team 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 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 × Female Investor=1	-0.003			0.590		
	(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.

Table 13: Experiment A 2nd Half Profile Evaluation Results (Gender, Contact-Based Heterogeneous Effect)

Dependent Variable	(1) Quality	(2) Collaboration	(3) Contact	(4) Investment
Panel A: $\beta^3 < 0$ (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
Panel B: $\beta^3 > 0$ (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
Investor FE	Yes	Yes	Yes	Yes
Observations	608	592	608	591

Notes. This table reports the contact decision-based heterogeneous effect of gender by using the second half of evaluation questions. Panel A reports the heterogeneous effect of investors who prefer not to contact female founders. Panel B reports the heterogeneous effect of investors who prefer to contact female founders. "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. 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 14: Experiment A 2nd Half Profile Evaluation Results (Race, Contact-Based Heterogeneous Effect)

Dependent Variable	(1) Quality	(2) Collaboration	(3) Contact	(4) Investment
Panel A: $\beta^3 < 0$ (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
Panel B: $\beta^3 > 0$ (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
Investor FE	Yes	Yes	Yes	Yes
Observations	608	592	608	591

Notes. This table reports the contact decision-based heterogeneous effect of race on the second half of the evaluation questions. Panel A reports the heterogeneous effect of investors who prefer not to contact Asian founders. Panel B reports the heterogeneous effect of investors who prefer to contact Asian founders. "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. 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 15: Experiment A 2nd Half Profile Evaluation Results (Age, Contact-Based Heterogeneous Effect)

Dependent Variable	(1) Quality	(2) Collaboration	(3) Contact	(4) Investment
Panel A: $\beta^3 < 0$ (Not Contact Asians)				
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
Panel B: $\beta^3 > 0$ (Contact Asians)				
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 age on investors' response to evaluation questions using the second half of profiles evaluated. Panel A reports the heterogeneous effect of investors who prefer not to contact older founders. Panel B reports the heterogeneous effect of investors who prefer to contact older founders. "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 and add the investor fixed effect. Standard errors in parentheses are bootstrapped for the two stage calculations. *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance at 1%, 5%, and 10% levels, respectively.

Figures

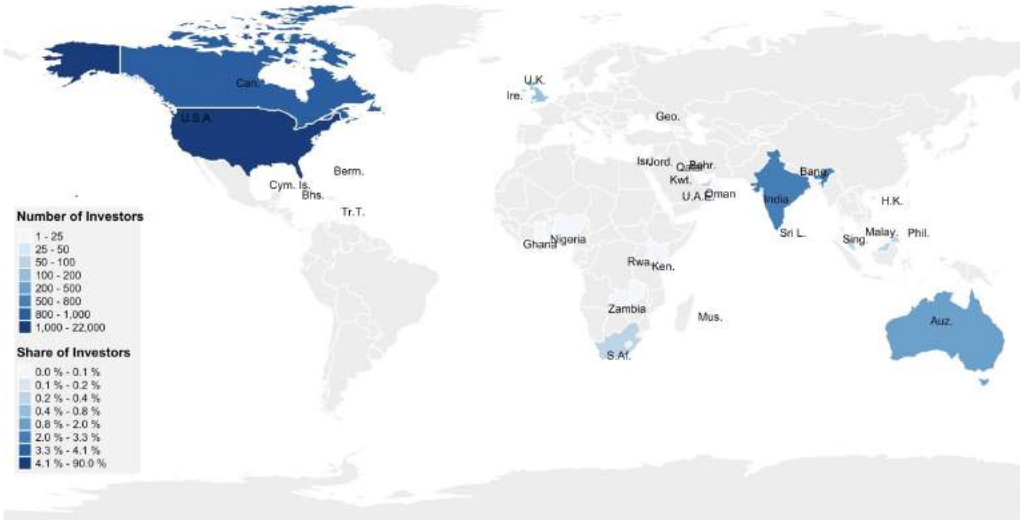


Figure 1: Geographical Distribution of Global Investors

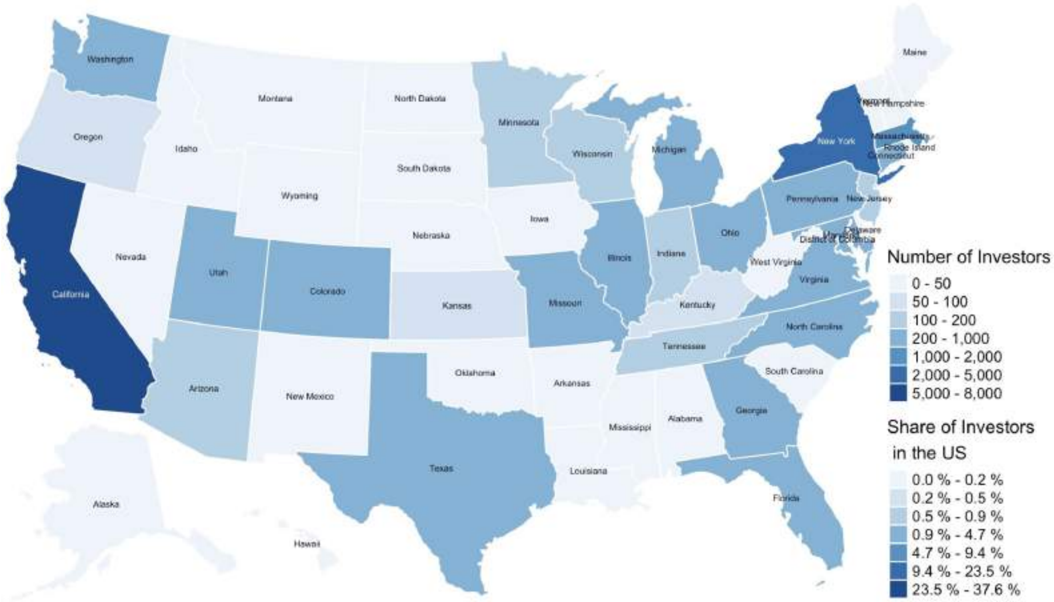


Figure 2: Geographical Distribution of U.S. Investors

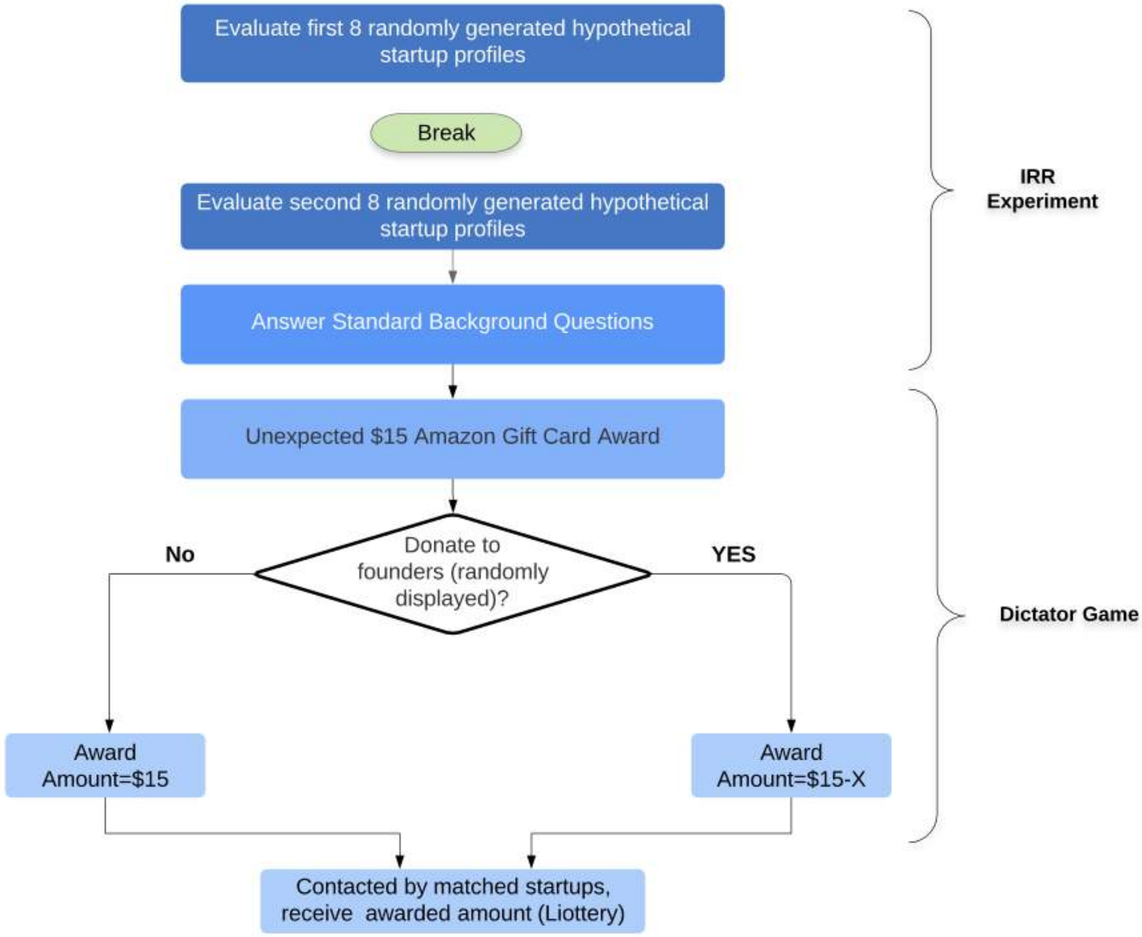


Figure 3: Experiment A Experimental Design

Figure 4: Founder Picture Example in the Donation Section

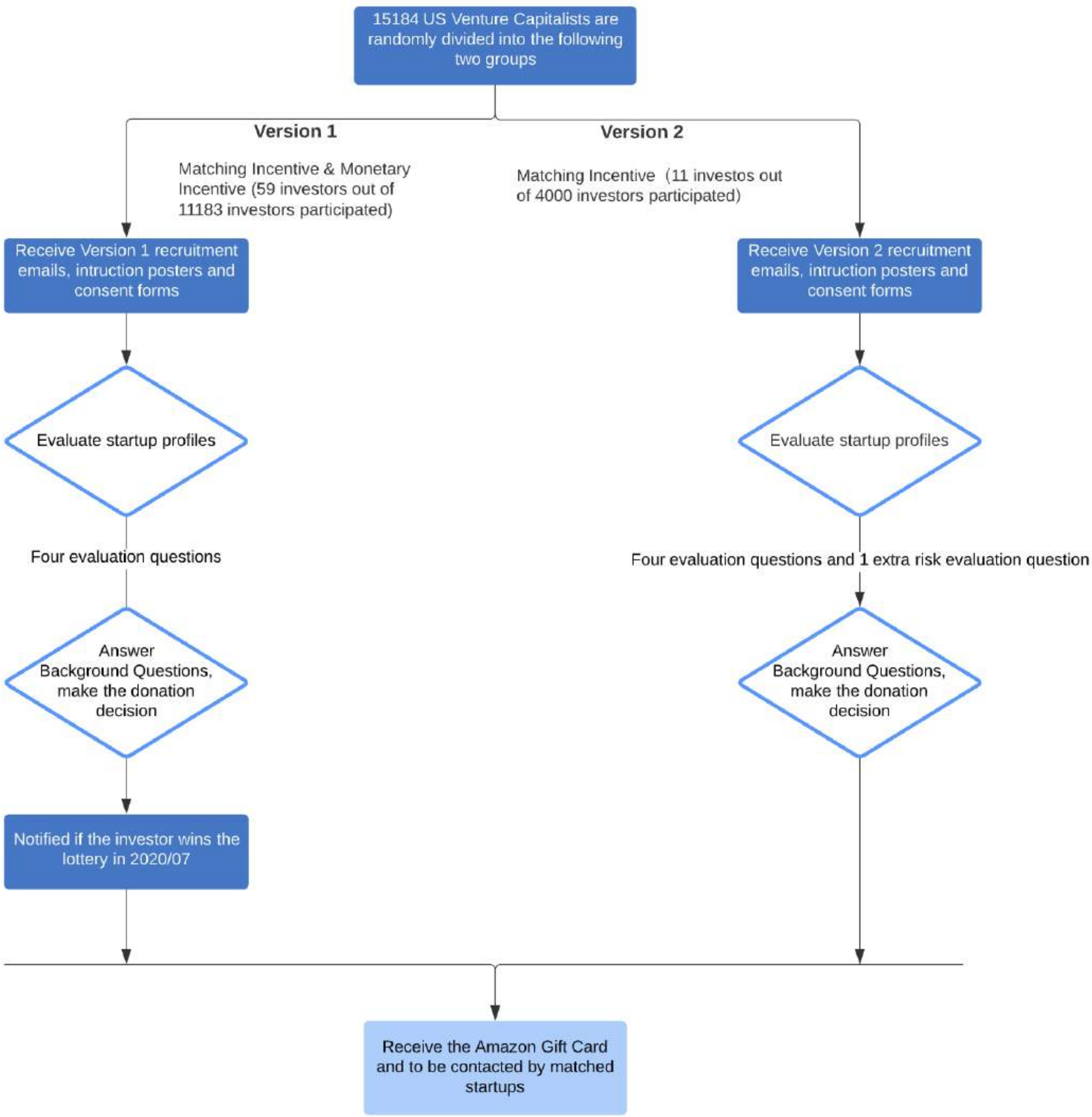


Figure 5: Experiment A Incentive Structure

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 6: Experiment B Example of a Pitch Email

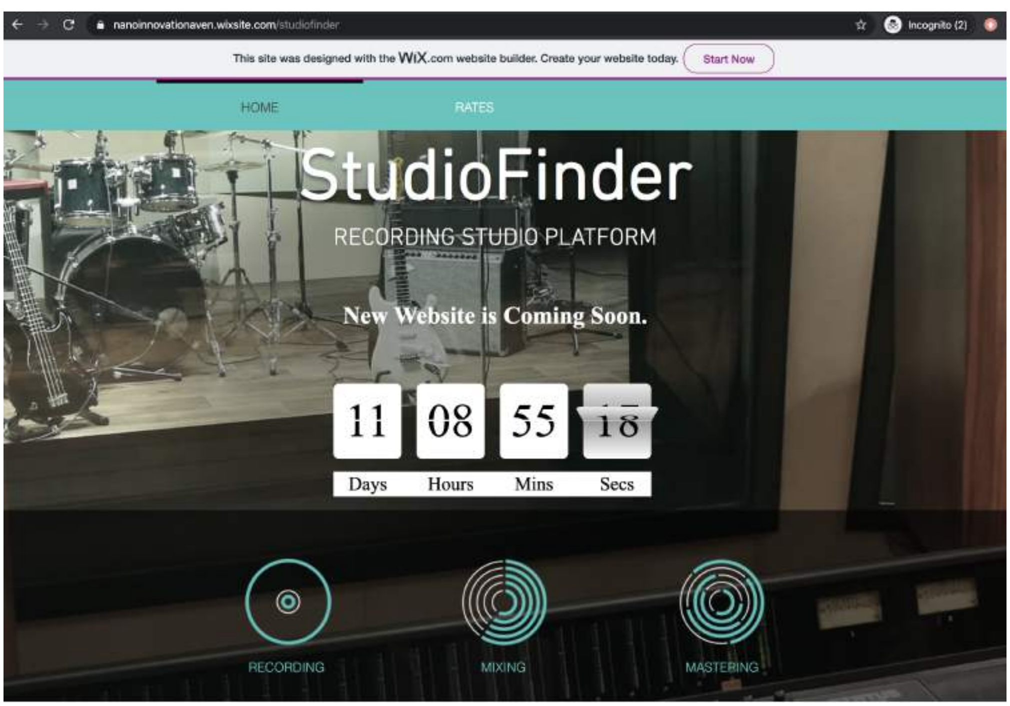


Figure 7: Experiment B Example of a Startup Website

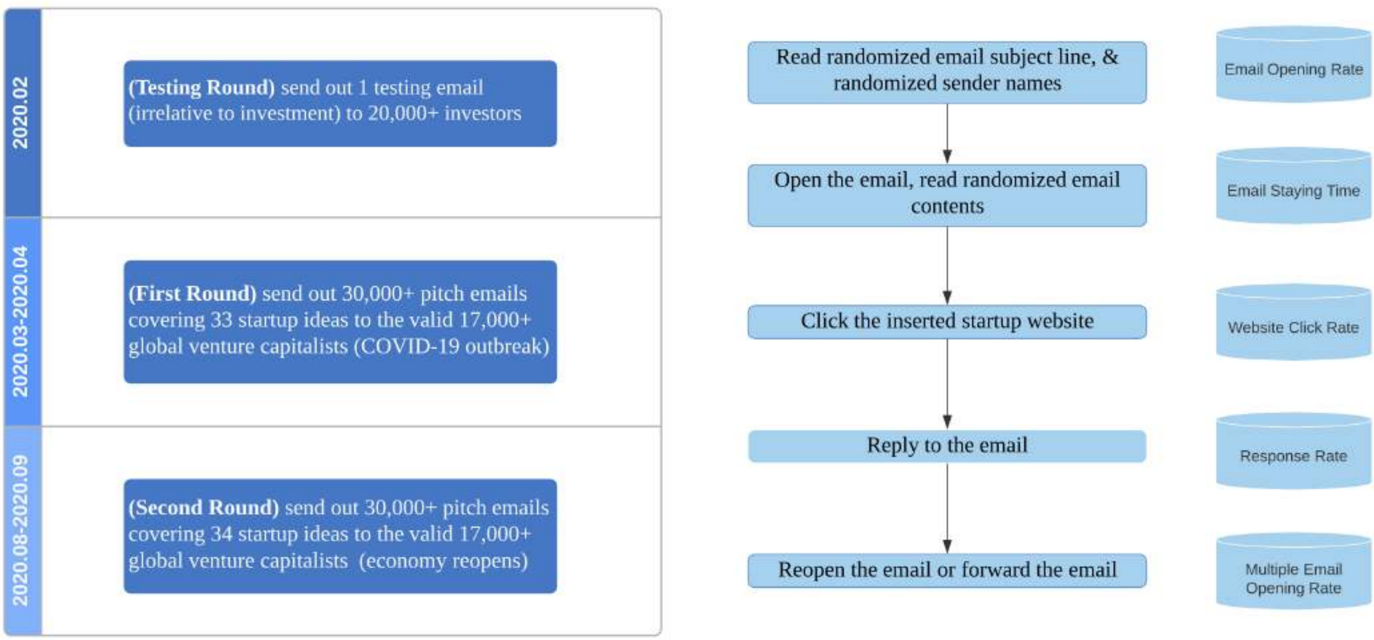


Figure 8: Experiment B Correspondence Test Experimental Design

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

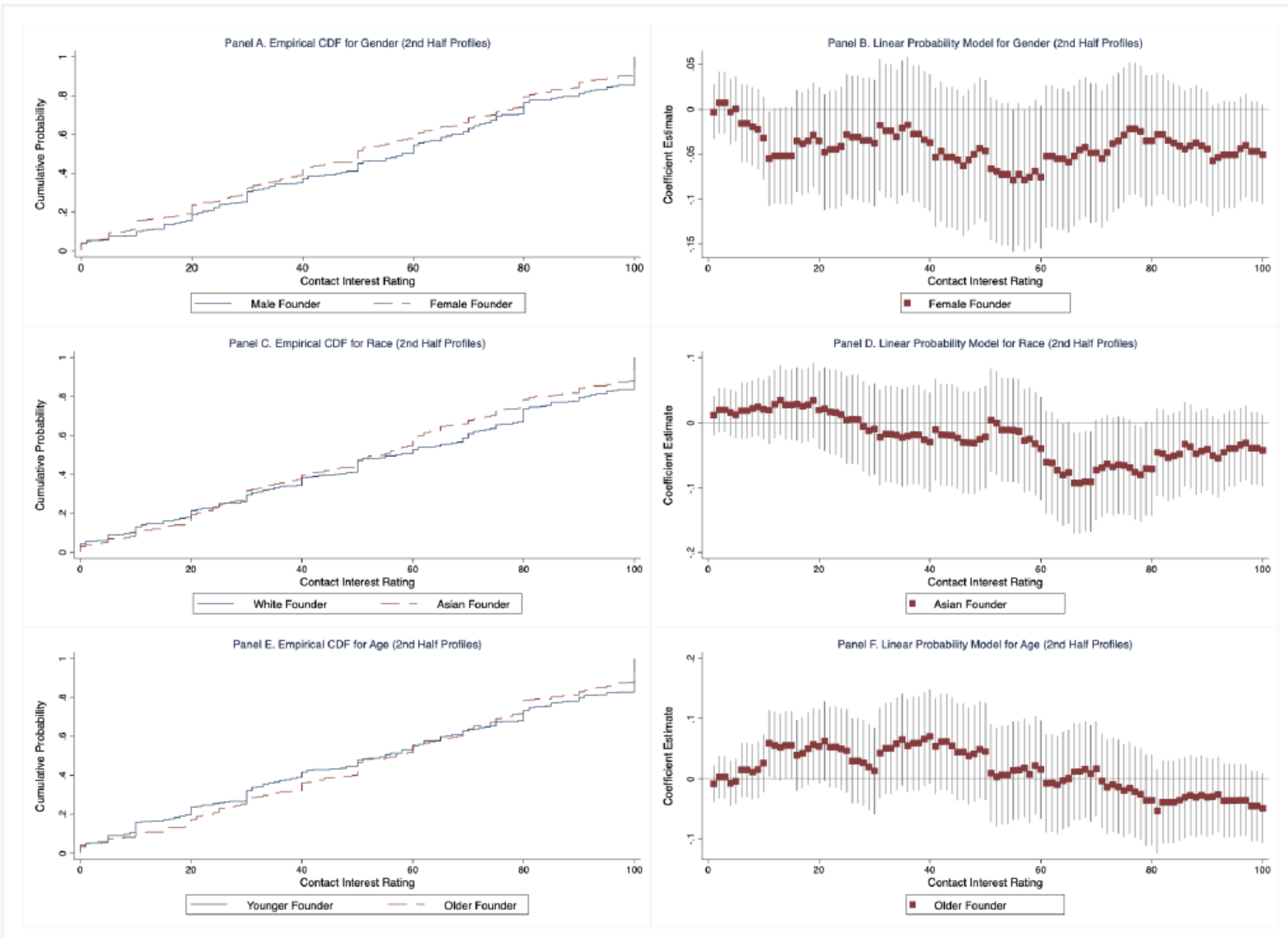


Figure 9: Effect of Founder's Gender, Race, and Age across the Contact Interest Distribution (2nd Half of Profiles)

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. $Pr(\text{Contact Interest} > x|\text{Female Founder})$ and $Pr(\text{Contact Interest} > x|\text{Male Founder})$). Panel B provides the OLS coefficient estimates (i.e. $Pr(\text{Contact Interest} > x|\text{Female Founder}) - Pr(\text{Contact Interest} > x|\text{Male Founder})$) 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.

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.

A.1.1 Pitchbook

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.

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.[107] 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.[108] It contains rich information about active Chinese 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,

[107] 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.
[108] Zdatabase description: http://www.p5w.net/fund/smjj/201209/P020120905327816063973.pdf

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.

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,[109] 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.[110]

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),[111] 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

[109]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.

[110]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%.

[111]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, we further checked these names to make sure that there are no names owned by famous startup founders or CEOs.

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 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 B5: 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.

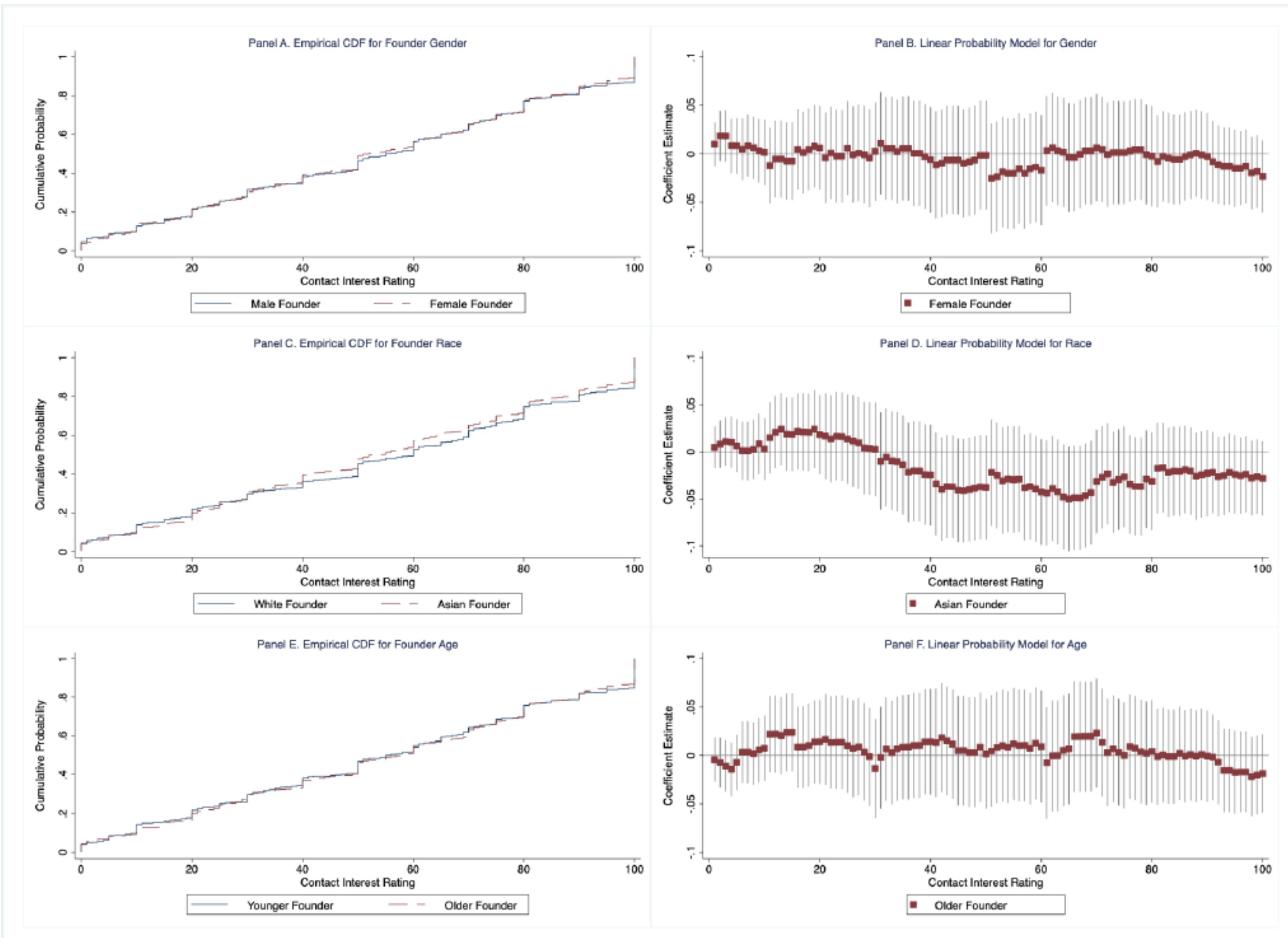


Figure B1: Effect of Founder's Gender, Race, and Age across the Contact Interest Distribution (Total Profiles)

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. $Pr(\text{Contact Interest} > x|\text{Female Founder})$ and $Pr(\text{Contact Interest} > x|\text{Male Founder})$). Panel B provides the OLS coefficient estimates (i.e. $Pr(\text{Contact Interest} > x|\text{Female Founder}) - Pr(\text{Contact Interest} > x|\text{Male Founder})$) 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.

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 B2: Experiment A Instruction Page (Version 2)

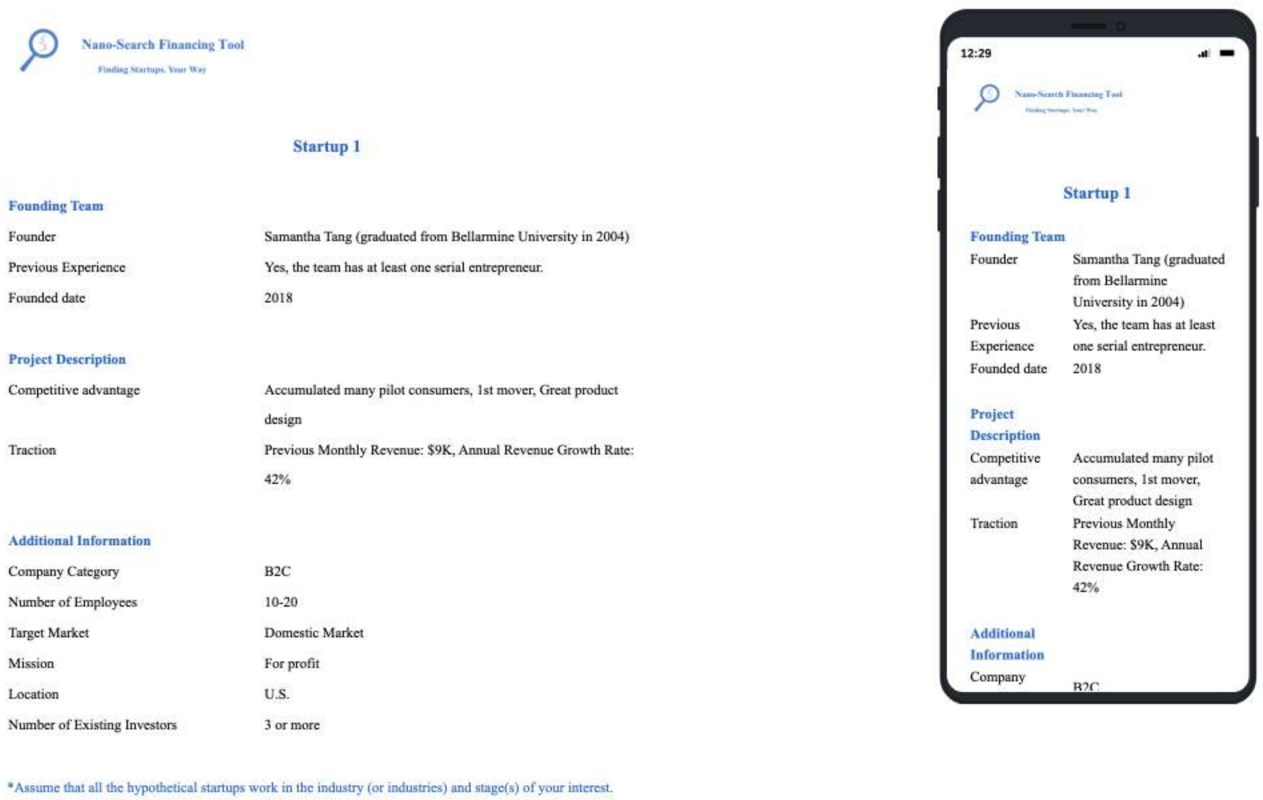
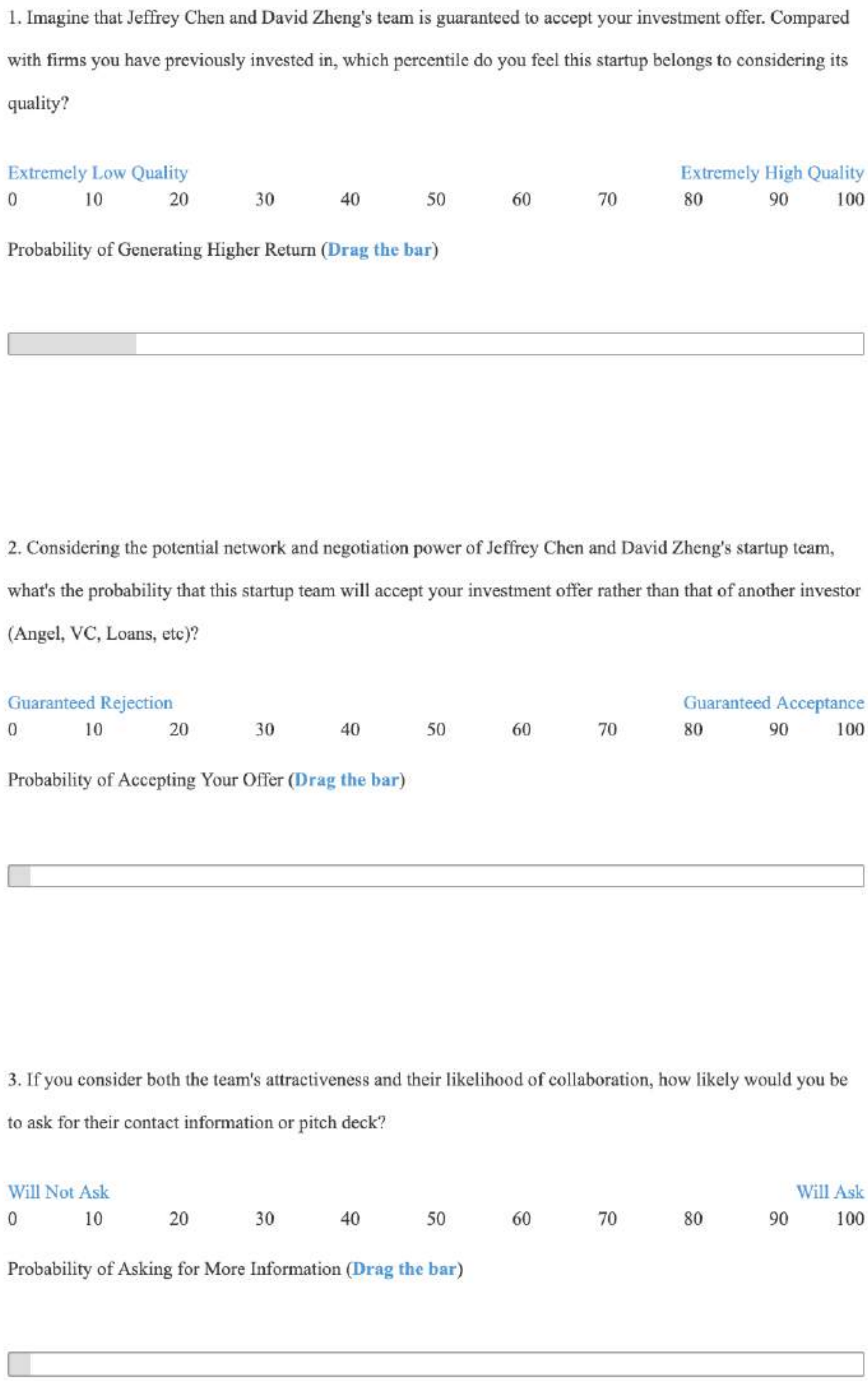


Figure B3: Experiment A Randomly Generated Startup Profile



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**)

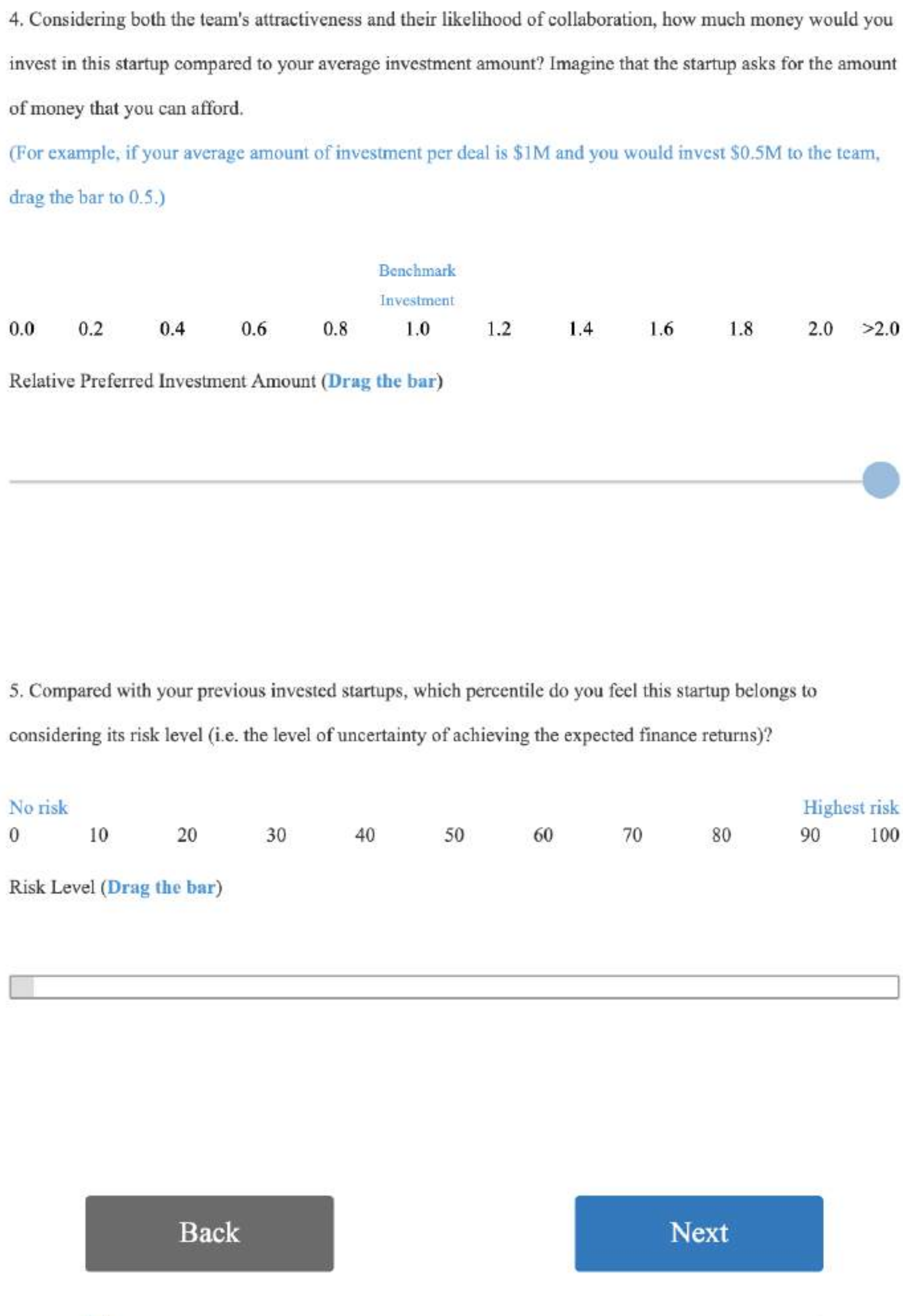
3. If you consider both the team's attractiveness and their likelihood of collaboration, how likely would you be to ask for their contact information or pitch deck?

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 B4: 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 B5: 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 B6: 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 B7: Experiment A Recruitment Email (Version 2)

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

Figure B8: 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.

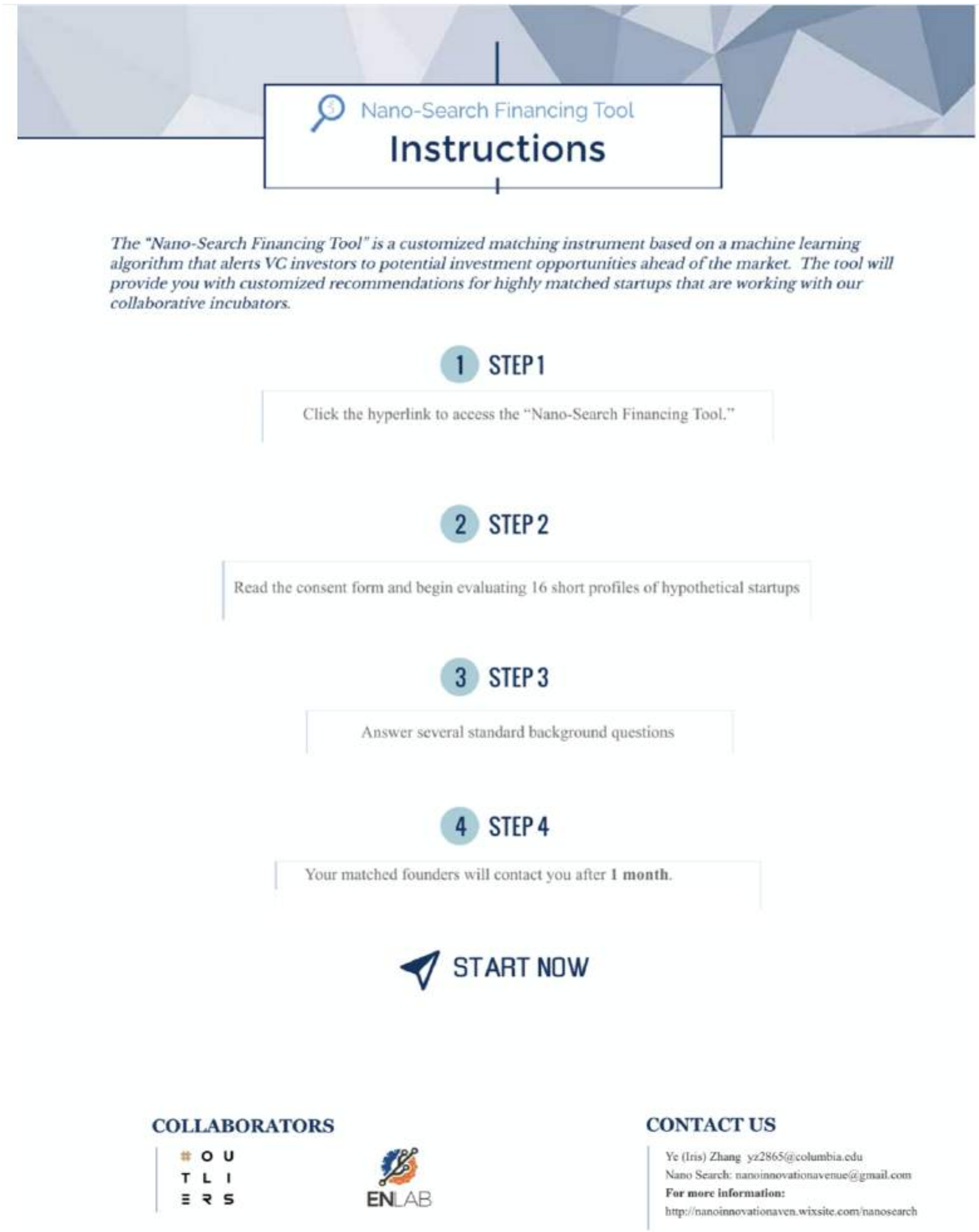


Figure B9: Experiment A Recruitment Poster (Version 2)

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

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,[112] birth records for selecting first names highly indicative of gender,[113] and 2010 U.S. Census data for generating last names highly indicative of race.[114] 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.[115] Then I pick the most frequent 100 names for each gender for further checks.[116]

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).[117] 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".[118]

Additional Check:
I also hire 107 Amazon Mechanical Turk users in the U.S. to confirm that the perception of gender and race elicited

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

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

[114] 2010 Census surnames product: https://www.census.gov/topics/population/genealogy/data/2010_surnames.html accessed on July 27, 2019.

[115] 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.

[116] 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.

[117] 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.

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

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.

---

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.

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)	(2)	(3)	(4)	(5)	(6)
	Full	Impact Fund	Common Fund	Full	Impact Funds	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)	(2)	(3)	(4)	(5)	(6)
	Full	Impact Fund	Common Fund	Full	Impact Funds	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, $\beta_{Asian} \neq 0$	$\times$	$\sigma^{II}_{AR} \neq 1$
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^{II}_{AR} 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.[119]

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

(spurious evidence of discrimination in favor of women)

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

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.)

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

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.

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

(*) 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:$\frac{(-c+\gamma)}{\underbrace{exp(\omega)}_{=\sigma_{FR}^{II}}}$ and $\frac{\beta_I}{\underbrace{exp(\omega)}_{=\sigma_{FR}^{II}}}$
The ratio of β_I and $\frac{\beta_I}{\underbrace{exp(\omega)}_{=\sigma_{FR}^{II}}}$ can identify $exp(\omega)$, which is equal to σ_{FR}^{II}.
With c and $exp(\omega)$, the expression of $\frac{(-c+\gamma)}{\underbrace{exp(\omega)}_{=\sigma_{FR}^{II}}}$ 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.[120]

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:

$$\begin{aligned} T_{ij} &= \quad 1\{c_1' < \beta X_1^{I*} + X_2^{II} + \gamma' G + \epsilon_{ij} < c_2'\} \\ T_{ij} &= \quad 1\{(c_1' - X_1^{I*} - \gamma' G)/\sigma_B < X_2^{II} + \epsilon_{ij} < (c_2' - X_1^{I*} - \gamma' G)/\sigma_B\} \end{aligned}$$

$$\prod_{i=1}^{n}(\Phi(\frac{(c_2' - X_1^{I*} - \gamma')}{\sigma_B^F}) - \Phi(\frac{(c_1'-X_1^{I*}-\gamma')}{\sigma_B^F}))^{T_{i\in F,j}=1}(\Phi(\frac{(c_2'-X_1^{I*})}{\sigma_B^M}) - \Phi(\frac{(c_1'-X_1^{I*})}{\sigma_B^M}))^{T_{i\in M,j}=1}$$

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

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))

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: $\beta_{ij}^{\hat{L}(1)} = \frac{\sum_{j' \neq j} X_{ij'} Y_{ij'}^{(1)}}{\sum_{j' \neq j} X_{ij'}^2}$ (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.