



**HAL**  
open science

# Crowd and the Gender Gap in Project Evaluations

Aurelien Quignon

► **To cite this version:**

| Aurelien Quignon. Crowd and the Gender Gap in Project Evaluations. 2024. hal-04792570

**HAL Id: hal-04792570**

**<https://hal.science/hal-04792570v1>**

Preprint submitted on 20 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Crowd and the Gender Gap in Project Evaluations

Aurelien Quignon\*<sup>1</sup>

<sup>1</sup>ICN Business School

<sup>2</sup>CEREFIGE, Université de Lorraine

## Abstract

Does relying on the crowd mitigate differences in evaluations between male- and female-led entrepreneurial projects? Drawing on the wisdom of the crowd and the gender stereotype theory, I examine gender differences in crowd evaluations and the role of gender identity in assessing quality using data from an online platform. I find that female-led projects do not receive lower evaluation scores, but female evaluators are more supportive of female founders. However, a gender gap persists across sub-criteria, with male-led projects receiving higher scores on traits associated with masculine stereotypes, while female-led projects garner more support for traits aligned with feminine stereotypes.

**Keywords:** Gender gap; crowd; idea evaluation

## 1 Introduction

Women tend to be underrepresented in many settings, and entrepreneurship is no exception. This underrepresentation has led to various public initiatives aimed at enhancing the financial inclusion and equality of women.<sup>1</sup> For instance, women account for only 14% of entrepreneurs in France. Similar patterns have emerged in the U.S., where female-led ventures represented just 6.8% of all deals in 2022 (Pitchbook, 2022). Several factors may explain the low share of women in entrepreneurship. This underrepresentation in high-growth entrepreneurship could

---

\*aurelien.quignon@icn-artem.com

<sup>1</sup>For instance, several initiatives have been introduced in France, at both national and regional levels, to support women's high growth entrepreneurship. They range from providing support to facilitate access to bank credit and support networks for women, such as Wom'energy.

be attributed to traditional factors such as gender differences in social capital, risk attitudes, and growth ambitions (Carter and Rosa, 1998; Cook et al., 2020). Discrimination and gender stereotypes against women may also explain the gender gap observed in venture screening (Guzman and Kacperczyk, 2019). Indeed, ventures founded by women tend to be judged less favorably than others across various entrepreneurial settings (Scott and Shu, 2017).<sup>2</sup>

Despite substantial evidence of gender biases in venture quality evaluations, our understanding of the gender gap in entrepreneurial assessments remains limited. Biases have been widely documented across various funding contexts, often involving a small group of professional stakeholders, such as early-stage investors (Ewens and Townsend, 2020) and accelerator programs (Dutt and Kaplan, 2018). More recently, a few studies have also investigated the effects of crowdfunding on gender gap (Bapna and Ganco, 2021; Greenberg and Mollick, 2017; Johnson et al., 2018), providing mixed results. Comparatively, little is known about how project quality assessments by non-expert stakeholders influence these disparities at other stages of entrepreneurship (Aldrich, 1999; Haynie et al., 2009).

In recent years, online platforms providing decentralized review systems have flourished (Quignon, 2023).<sup>3</sup> In crowdsourcing contexts, user ratings serve as a proxy for idea quality, offering valuable insights for founders (Riedl et al., 2010). Leveraging a general audience in an online context may benefit founders at every stage of development by changing the composition and number of stakeholders involved in idea evaluation (Bapna and Ganco, 2021). By relying on diverse preferences from a broader audience of non-experts—the wisdom of the crowd—the evaluation of female-led ventures may be influenced by the characteristics of the crowd, such as gender and preferences for innovation. As a result, decision-making regarding idea quality can move away from gender as a heuristic and remove bias from idea evaluation (Cumming et al., 2019; Surowiecki, 2005). Yet, few studies have examined the role of online platforms that leverage crowds in reducing the gender gap in evaluations.

Understanding whether crowd evaluators are gender-neutral in their assessments of entrepreneurial ideas has important implications for founders, investors, and other stakeholders in entrepreneurship. Early-stage evaluations of business ideas are especially critical, as positive

---

<sup>2</sup>See, for instance, Bagues et al. (2017), Li (2017), and Li and Agha (2015) how are biased evaluations concerning the gender of candidates in academia and research funding.

<sup>3</sup>Product-oriented platforms have popularized crowd evaluations, such as Amazon and Yelp (Reimers and Waldfogel, 2021). Crowds in screening and selecting ideas are increasingly used in different settings, including product implementation (Dahlander et al., 2023b; Hoornaert et al., 2017), creative industries (Luo et al., 2021), and science (Franzoni and Sauermann, 2014).

or negative assessments can significantly influence the entrepreneurial process and founders' pursuit of their ideas (Gupta et al., 2014). Analyzing how evaluators score business ideas across multiple criteria is essential for uncovering the mechanisms that drive overall ratings and shape founders' beliefs about their ideas' potential. Furthermore, evaluations in low-stakes decision-making contexts provide a cost-effective means for idea assessment, compared to other settings. Addressing these questions is crucial for enhancing the pipeline of women pursuing entrepreneurial opportunities and transitioning into high-growth ventures.

In this paper, I examine crowd evaluations as a remedy to narrow the gender gap in the evaluation of early-stage projects. Drawing on the wisdom of the crowd (Surowiecki, 2005) and gender stereotypical beliefs (Eagly and Karau, 2002; Yang et al., 2020), I argue that incorporating diverse opinions from various stakeholders can mitigate gender disparities in the evaluation of entrepreneurial ideas. However, founders are likely to receive different evaluations depending on whether sub-criteria are aligned with stereotypical gender attributes. I use data from an online evaluation platform where founders apply for feedback in the form of ratings from a general audience. This setting includes several features. First, the platform is open and free of charge (i.e., low-stake setting) to anyone who wishes to discover more about the ecosystem and give their opinions. This allows me to evaluate decision-making from different stakeholders, observing individual ratings and whether the evaluator expresses interest in the project. Furthermore, evaluators can observe all projects with the same level of information. Second, the platform includes projects with varying objectives, ranging from seeking feedback on their ideas to fundraising. This setting has the advantage of examining the gender gap at different stages, reducing concerns about studying decision-making based on a sample of funded ventures where women are underrepresented. Finally, I observe the set of early-stage projects submitted on the platform. This set not only contains information on the gender of the founding team but also detailed evaluations, sub-criteria, investment recommendations, and evaluator gender for each project. This information is augmented by project characteristics to measure quality. The data includes 407 early-stage projects from 2015 to 2018 and evaluated by 2,226 evaluators.

The results can be summarized as follows: I first show that female-led ventures receive lower project evaluations than their male counterparts. The estimated gender gap is equal to 0.08 standard deviations (SD) in the evaluation scores from the crowd after controlling for evaluator fixed effects. In other words, female founders received on average a 0.61 lower rat-

ing than male founders. After conditioning on projects' observable characteristics available to evaluators on the platform, I do not observe a gender gap in evaluation scores. The absence of a gender gap is confirmed when examining whether evaluators express differential investment interests. I find no significant difference, suggesting that leveraging diverse opinions can narrow the gender gap in evaluations. These results are qualitatively similar when using an entropy matching reweighting procedure to balance the observable characteristics of female- and male-led ventures or the gender composition of evaluators. Additionally, to further mitigate concerns about residual bias from unobservable factors, I control for extensive project characteristics and estimate the [Oster \(2019\)](#) bounds, thereby supporting that omitted variables do not drive the relationship between gender and evaluation outcomes.

I then extend the analysis by supplementing the results from the aggregate ratings with an examination of the sub-criteria available in the data, exploring how gender stereotypes contribute to the observed absence of an overall gender gap. I find that aggregate estimates hide substantial heterogeneity. Female-led ventures are perceived as less likely to succeed, less ambitious, and less feasible than their counterparts, with an estimated gap ranging from 0.08 to 0.19 SD. In contrast, evaluators give higher scores for environmental and societal criteria for female-led ventures. These results suggest that, despite the large number of opinions, gender stereotypes persist in specific criteria within crowd evaluations.

Using the approach of [Ahrensjö et al. \(2022\)](#), I also explore the role of gender identity in evaluator decision-making. This approach allows me to test the importance of a larger proportion of evaluators who share the same gender in terms of both the rating and investment interest. I find no evidence that a crowd composed of a larger share of stakeholders of the same gender as the founding teams significantly benefits one of the two groups. However, female-led projects received significantly more favorable evaluations of the idea's quality if the crowd was composed only of women, compared to male-led projects evaluated by a crowd that matched their gender. As a result, this suggests that favoritism is concentrated in female groups.

This paper makes main contributions to several strands of the literature. First, I contribute to the literature on entrepreneurial idea evaluation, which has received limited attention ([Gupta et al., 2014](#); [Mitchell and Shepherd, 2010](#)), by integrating the wisdom of the crowd in a nonpecuniary context. The results provide compelling evidence that leveraging opinions from various stakeholders can narrow the gender differences in evaluation and can be effective in reducing the underrepresentation of female participants. Furthermore, this paper goes beyond

considering gender differences in evaluation within a single sector and encompasses projects from various sectors, ranging from low- to high-tech intensive. I also contribute to the literature on gender stereotypes, which typically examines how gender influences decision-making through social norms. Previous studies have shown that women are often disadvantaged in decision-making processes, such as fundraising, due to perceptions shaped by entrepreneurial and social attributes of ventures (Johnson et al., 2018; Lee and Huang, 2018; Yang et al., 2020). I extend this body of work by showing that focusing on a single criterion (e.g., success) hides important heterogeneity in the underlying factors used by evaluators to assess project quality. By demonstrating that evaluators are more supportive of male-led projects based on economic expectations and female-led projects based on social expectations, this paper contributes to a more nuanced understanding of the role of gender stereotypes on evaluation. More broadly, this paper is also related to the growing literature on gender bias in online context. Prior research highlights behavioral differences in the gender of evaluators in the case of product evaluations (Bayerl et al., 2024; Proserpio et al., 2021). The results suggest that female evaluators react differently to founder gender, being more supportive of women while men are neutral.

The rest of the paper is organized as follows. Section 2 presents the conceptual background. Section 3 details the platform and data. Section 4 presents the empirical strategy to estimate the gender gap in idea evaluations. Section 5 presents the results and robustness checks, and Section 6 concludes.

## **2 Conceptual Background**

### **2.1 Gender Bias in Project Evaluation**

Similar to the innovation process that results in novel ideas, entrepreneurship usually involves several steps, including creation, evaluation, selection, execution, and financing (Guzman and Kacperczyk, 2019). Judging the quality of novel business ideas is an important step that leads to subsequent opportunities (Haynie et al., 2009; Mitchell and Shepherd, 2010). In particular, stakeholder decision-making allows ventures to participate in selective programs and receive essential financing resources to reach the next stage of development (Lee and Huang, 2018). However, evaluating early-stage idea quality is difficult for a wide range of stakeholders, from non-experts to professionals. The uncertainty surrounding technology, market potential, and the founding team is exacerbated in the early stages by limited information and

unpredictable outcomes. The attributes by which stakeholders scrutinize and select new entrepreneurial projects have been extensively examined (Bernstein and Korteweg, 2017; Scott et al., 2020), encompassing factors like level of education, previous successful crowdfunding campaigns, and the founding team (Zhang, 2009). Business plan competitions, incubator and accelerator programs, and crowdfunding campaigns that evaluate early-stage concepts may lack financial performance data (i.e., hard data) or demonstration models. This is consistent with recent evidence that many ventures fail, even those that successfully raised funds (Hall and Woodward, 2010). As a result, evaluators are unable to assess the quality of business ideas accurately and must rely on subjective signals to form beliefs about the high-growth potential of nascent ventures. Subjective signals may be based on several pieces of information unrelated to the quality of early-stage ventures, such as communication style, the order of presentations during pitching, and the main characteristics of the founders. Among these characteristics, the gender of the entrepreneur is the most observable to evaluators (Kanze et al., 2020). According to Bernstein and Korteweg (2017), investors' decision-making on AngelList is most adversely affected when details about the founding team are omitted.

An extensive body of literature documents that female-led startups are judged more negatively than male-led startups, even when they have comparable business ideas and credentials (Lee and Huang, 2018). Prior research at the financing stage suggests that angel investors (Brooks et al., 2014; Ewens and Townsend, 2020) and venture capitalists (Brush et al., 2003) may have a negative opinion of the quality of new ventures led by women. Consequently, female-led startups are more likely to be perceived as less successful than their male counterparts, leading to resource constraints and a gender gap in high-growth entrepreneurship (Ewens and Townsend, 2020; Lee and Huang, 2018). Several explanations have been proposed, such as differences in social networks.<sup>4</sup> However, gender stereotypes are more prevalent in evaluations when assessing funding (Guzman and Kacperczyk, 2019; Lee and Huang, 2018), as men are perceived as more successful (Aldrich and Yang, 2013). According to the stereotype mechanism, early-stage investors' perceptions of female entrepreneurs' overall fit and aptitude are misaligned. As a result, women tend to be perceived as less competent than their counterparts and less likely to lead entrepreneurial projects to success. Brooks et al. (2014) demonstrate that entrepreneurial projects pitched with a male voice are more likely to receive investment interest

---

<sup>4</sup>Social networks are an important resource for entrepreneurs that enable them to reach mentors and potential investors. However, female entrepreneurs tend to have different networks than their male counterparts, which prevents them from accessing valuable information and opportunities to start new ventures.

than those pitched with a female voice. In other experimental studies, [Thébaud \(2015\)](#) finds that students are biased against women who present entrepreneurial ideas, even when the business pitch is kept constant. These negative stereotypes can be exacerbated by inexperienced evaluators ([Bapna, 2019](#)) or in settings where information is lacking ([Simcoe and Waguespack, 2011](#)), increasing reliance on gender-based heuristics when evaluating the potential of early-stage projects. For instance, [Bapna and Ganco \(2021\)](#) find that less experienced female investors tend to favor female-led startups, whereas this preference is not observed among experienced female investors.

Finally, gender homophily may contribute to the observed gender gap in entrepreneurship. Investors tend to base their decisions on in-group preferences for founders with similar traits, such as gender. In other words, early-stage investors may have greater confidence when evaluating the quality of ideas led by female founders ([Ewens and Townsend, 2020](#); [Snellman and Solal, 2023](#)). The influence of homophily should be particularly harmful to early-stage female-led startups because the funding stage is a male-dominated industry with only 6% of VC partners and 20% of angel investors being female ([Brush et al., 2014](#); [Sohl, 2018](#)). Taken together, these findings suggest that female-led startups are perceived as less competent in the entrepreneurship context.

## 2.2 Project Evaluation through the Crowd

Negative bias against women is not limited to the funding stage. The underrepresentation of women might occur at earlier stages of the process, such as business evaluations ([Guzman and Kacperczyk, 2019](#)). Several mechanisms have been introduced to mitigate gender-based discrimination in evaluation and selection, such as blinding to ascriptive cues. For instance, [Goldin and Rouse \(2000\)](#) show a higher likelihood of women being selected through blind intervention. Similarly, ([Dahlander et al., 2023b](#)) find that gender bias was removed in blinded idea evaluation. However, [Kolev et al. \(2019\)](#) show that women received lower evaluation scores in blinded reviews, suggesting that this is not a sufficient tool for reducing bias. In addition, this mechanism can be time-consuming to implement on a large scale and create opportunity costs ([Dahlander et al., 2023b](#)).

In the context of project screening, an online platform that allows different stakeholders to assess the quality of new ventures is relevant. Involving crowds in project selection using online platforms is increasingly common in R&D and entrepreneurial projects ([Dahlander et al.,](#)



2023a; Quignon, 2023). Receiving crowd evaluations about the underlying potential of business ideas can reduce the gender gap, enabling an increase in the number of new ventures led by women. Crowdsourcing evaluation is an online distributed system that allows a large and geographically dispersed audience of stakeholders ranging from non-experts to experts (i.e., angels and VCs) to participate. In addition, early-stage founders are not required to pitch ideas in front of evaluators because decision-making is based on information disclosed on the platform. By rating the projects, the crowd expresses their belief about the potential based on the information disclosed by the team, which may help them refine their concepts (Riedl et al., 2013). Recent empirical evidence supports the entrepreneurial learning process in crowdsourcing feedback (Quignon, 2023).

Based on the wisdom of the crowd mechanism (Galton, 1907; Surowiecki, 2005), crowds can mitigate gender stereotypes in project evaluations and ultimately increase the female entrepreneurship rate. Online platforms enable early-stage ventures to leverage a wider audience with diverse experiences, ranging from professional investors to amateurs or customers (Mollick and Nanda, 2016). Therefore, evaluations that traditionally relied on a small set of experts shifted to diverse and different knowledge groups, enabling a change in the composition and number of potential evaluators for new ideas (Davidsson et al., 2021). This is reinforced in non-pecuniary settings that do not impose personal resources for evaluating and recommending early-stage ventures, thus reducing concerns about the selection of wealthy individuals (Sauer-mann et al., 2020).<sup>5</sup> Thus, the collection of more varied opinions is essential to the wisdom of the crowd, in which evaluators make different types of judgment mistakes that cancel each other out when judgments are aggregated. As a result, the average judgment error tends to be less than the error of an individual expert, which limits the effect of status (Müller-Trede et al., 2018). These arguments are consistent with recent evidence that, aside from inexperienced female investors, crowd backers' assessment of entrepreneurial ideas does not discriminate against female-led ventures (Bapna and Ganco, 2021). Decision-making is motivated by the main characteristics of the product or service proposed and the motivation and engagement of the founding team, rather than ascriptive factors such as gender (Shafi, 2019). However, studies have suggested that crowd evaluations generally agree with experts when selecting promising entrepreneurial ideas (Magnusson et al., 2014; Mollick and Nanda, 2016), which can replicate

---

<sup>5</sup>Low-stakes crowdsourcing ideas are popular since they involve small donations, such as reward-based crowd-funding. For instance, the most popular amount from the crowd is \$25 on Kickstarter (Bapna and Ganco, 2021). In other settings, the crowd is only solicited to rate ideas, such as t-shirts (Lakhani, 2016) or logo designs (Gross, 2020).

the decision-making of professional investors based on gender heuristics. Therefore, I argue that crowd evaluation of entrepreneurial projects could be an attractive and useful complement to the expert selection process in low-stakes scenarios, helping to mitigate discrimination against women and enhance female participation.

The wisdom of the crowd suggests that aggregating anonymous opinions should reduce differences in evaluation scores between male and female founders by giving more weight to diverse characteristics. Given that evaluations of early-stage ideas rely on the founders' perception of their aptitude to develop their ideas into successful ventures, the underlying criteria forming the aggregate score can suffer from stereotypes or cognitive biases (Huang and Pearce, 2015; Reskin, 2003). Stereotypical beliefs—such as those rooted in social norms—may still influence evaluators' perceptions of project quality and affect how they rate specific sub-criteria on which aggregate scores are based (Abraham et al., 2024; Gupta et al., 2014). Stereotypical beliefs refer to cognitive schemas that individuals routinely use to represent differences among members of social groups, potentially perpetuating or mitigating evaluative inequalities (Bordalo et al., 2016; Hilton and von Hippel, 1996). These beliefs are correlated with attributes such as competence, ambition, risk-taking, and social commitment (Fiske et al., 2002), which are associated with either lower or higher social status (Yang et al., 2020). Furthermore, in male-dominated contexts, such beliefs are reinforced, creating expectations that men are stronger performers and are evaluated more favorably (Correll and Benard, 2006; Dardenne et al., 2007). For instance, Wu and Chua (2011) show that female founders are judged as less competent compared to their male counterparts. Similarly, Lee and Huang (2018) find that female founders are assessed as less viable, which is consistent with the stereotypical beliefs that competence and success, associated with ambition, is restricted to men (Eagly and Karau, 2002; Fiske et al., 2002). Conversely, women are expected to align with social norms related to societal and environmental involvement, such as caring and trustworthiness. For instance, Lee and Huang (2018) show that decision-makers perceive the social framing of ventures as a positive signal, reducing the gap across female- and male-led ventures. In the context of accelerators' selection, Yang et al. (2020) suggest that social credibility, acting as a signal, increase the likelihood of selection of women-led ventures because of the congruency with gender expectations. Given the prevalence of stereotypical beliefs influencing how evaluators perceive and assess the quality of early-stage entrepreneurial ideas, I expect the gender gap to persist across sub-criteria. Specifically, female-led projects are likely to receive lower evaluations on

criteria traditionally associated with men, such as success and ambition, while receiving higher evaluations on criteria aligned with stereotypically feminine attributes, such as affection, and social and environmental impact.

## **3 Context and Data**

### **3.1 Setting: Crowdrating Platform**

To examine whether founding team gender affects evaluation of early-stage projects, I take advantage of a French online platform—Wirate. The online platform allows founders to raise external feedback through ratings from an online community. The platform was launched in January 2015. The platform is free and publicly available to founders and evaluators. Between 2015 and 2018, 837 entrepreneurial projects were launched on the platform and 2,600 evaluators registered to help early-stage entrepreneurs develop their ventures by assessing the potential of their ideas.

Founders create a project page by providing a short description of the idea and additional information, such as a business plan and video. Like crowdfunding platforms, such as Kickstarter, projects registered on the platform are listed on a webpage for entrepreneurs and evaluators, with a short project description and a visual. Almost all the pages are in French (95%). Importantly, evaluators are not biased by prior evaluation scores because they are not visible to potential evaluators, except for the underlying number of ratings. This ensures that the wisdom of the crowd is not undermined. Based on published information, an evaluator can rate the project on a scale ranging from 0 (worst) to 10 (best). In addition, evaluators can post comments (i.e., feedback) on the project page in the form of questions, encouragement, and opinions, enabling them to assess the quality of the startups. The quality is evaluated using a series of seven predefined questions representing the evaluation criteria listed in Table 1. Thus, I focus on overall ratings and sub-ratings, which allows me to examine whether evaluators evaluate quality criteria differently.

The data were derived from information provided by Wirate. The sample consists of early-stage startups registered on the platform from 2015 to 2018. The information includes detailed characteristics of entrepreneurial projects: project name, description of the idea, registration

date, technological category<sup>6</sup>, rating, sub-criteria ranging from 0 to 10, and underlying number of evaluations. Importantly, once the project was launched on the platform, the pages of the project remained online with the relevant information, thus reducing survival bias (Yu, 2020). I merged these projects with information from various external sources. Specifically, I hand-collected information from Dealroom, Crunchbase, LinkedIn, and myfrenchstartup.com (see Appendix Table 7). Finally, evaluator-level information is extracted from the platform. For each evaluator, I collect information including the evaluator’s name, evaluation rating for the evaluated projects, the date on which they leave a comment, and the total number of evaluations submitted.

Given the research question, I implemented several sample restrictions. As a first sample restriction, I require that the project receive at least one evaluation during the sample period to capture the gender gap. Second, I restrict attention to startups that satisfy the condition of being founded by males or females only, excluding gender-mix founding teams. Finally, I drop startups with missing characteristics to, control venture quality (Yu, 2020). These sample selection criteria reduced the sample to 4,860 project-evaluator level observations, including 407 early-stage ventures and 2,226 evaluators.

Table 1. Rating Process

Rating (0-10)	Success	<i>"In your opinion, is this offer likely to succeed?"</i>
	Useful	<i>"Does this project seem useful?"</i>
	Originality	<i>"Do you think it's original?"</i>
	Ambition	<i>"Is it ambitious?"</i>
	Feasibility	<i>"Is it feasible?"</i>
	Societal and Environmental Impact	<i>"Does it have an ecological or societal dimension?"</i>
	Affection	<i>"Do you like this project?"</i>

### 3.2 Dependent Variables

The main dependent variable is a standardized score (i.e., z-score) derived from the overall rating (**Rating**) and the sub-ratings of success, usefulness, originality, ambition, feasibility, CSR, and affection on which it is based. I use standardized scores at the project-evaluator level to make the regression coefficients comparable. As a robustness check, I re-estimate the specifications with nominal ratings as the dependent variable. To test the extent to which evaluators penalize early-stage female-led startups in their quality assessment, I also consider

<sup>6</sup>Overall, founders can select a technological category from the detailed 38 sectors (items) proposed by the platform, grouped into 19 macro-sectors. For example, these sectors include fashion, arts, health, fintech, robotics, and energy.

whether an evaluator is willing to invest in a project (**Investment Interest**). This dependent variable is a dummy that equals one if a project receives investment interest from an evaluator. This information is derived from a final question that did not enter the rating computation. Both dependent variables constitute a non-pecuniary involvement of evaluators. This non-pecuniary feature helps attract a larger audience and reduces concerns about selecting high-income individuals (Hegde and Sampat, 2015).

### 3.3 Independent Variables

The main independent variable is a dummy equal to one if the founding team is composed exclusively of female, and zero otherwise (**Female**). I identify the gender of each member of the founding team based on the first names provided in the description of the project or business plan. Specifically, I used the API Genderize.io. (Santamaría and Mihaljević, 2018), which predicts gender based on first names with an associated probability. One advantage of the sample is that most names are French, thereby limiting country-specific bias in gender imputation. To increase accuracy, I hand-checked the results and used pictures when available, allowing me to assign gender to all members of the founding team (i.e., the CEO and other members of the management team) in the analysis sample. As an additional robustness check, I perform the same analysis including mixed-gender teams.

To further explore gender bias in project evaluations, I test several empirical specifications with interaction terms. Specifically, the main variable of interest is interact with the a categorical variable for the development stage of the project. When registering, founders must select the development stage of their early-stage entrepreneurial project based on five predefined categories: **Idea**; **Development**; **Incubation**; **Product Market Fit**; and **Launched**. These categories reflect the maturity (i.e., uncertainty) of projects and can also improve evaluators' assessment based on early proof of success.

Finally, to investigate how the gender of evaluators impacts the gender gap, I take advantage of the information on the first name of each evaluator to identify their gender in a similar vein to that of the founding team. Then, I follow Ahrsjö et al. (2022) and construct an indicator of **Gender-Identity**. This indicator measures the extent to which evaluators align with the gender composition of the project founders (female and male) by counting the number of evaluators who share the same gender as the founding team. For instance, **Gender-Identity** is equal to zero if a project composed of women received only evaluations from men and one if

all evaluations emanated from women.

### 3.4 Control Variables for Project Quality

Early-stage startups vary widely in quality and the probability of high-growth realizations, such as Initial Public Offering (IPO) or acquisitions (Guzman and Kacperczyk, 2019). Therefore, several variables are included in the regressions to control for differences in project quality that could confound the relationship between founders' gender and crowd evaluations. I control for the project's main characteristics by including a set of variables measured at registration or one year before the evaluation process. First, I include the size of the founding team (**N. Entrepreneurs**) involved in the project based on the information available on the platform. This variable signals human and social capital that can be associated with higher skills for identifying new business opportunities and realizing a development strategy (Smith et al., 2001; Snellman and Solal, 2023). I consider **Venture Age**, which is defined as the time frame between the platform's release date and the year the project began (e.g., incorporation date or year declared on pitch/LinkedIn profiles). Older ventures experience more time developing their business idea. The venture age is also related to success in crowdfunding campaigns (Prokop and Wang, 2021). I also introduce a dummy variable to determine whether a venture is incorporated into a commerce registry (**Incorporated**), whether a venture has participated in incubator or accelerator programs (**Incubator/Accelerator**). I use a dummy variable equal to one if ventures have already raised funds from professional investors, such as angels or VCs, and zero otherwise (**Prior Funding**).<sup>7</sup> This variable captures project quality as assessed by professional investors, thus providing valuable third-party information to evaluators. The dummy variable for the foreign status of the project (**Foreign Venture**) is also introduced to control for potential bias from evaluators (Wright et al., 2023), because the location, including city and country, is apparent on each project page.

Following Bapna and Ganco (2021), I include in a robustness check, a dummy variable to determine whether an evaluator has previously assessed a project (**Previous Evaluation**). Early-stage projects are uncertain, resulting in evaluators relying on gender heuristics as a signal of quality. This could be more severe for first-time evaluators who, are more likely to use this characteristic in their assessments (Bapna and Ganco, 2021; Bernstein and Korteweg,

---

<sup>7</sup>It is noteworthy that I included a dummy variable rather than the amount raised because financing events account for about 18% of the ventures, with a lot of amounts that are unavailable. Among this subsample, I retrieve the level of financing for about 10%.

2017). In contrast, experienced evaluators can differentiate project quality using several characteristics other than gender.

Finally, I include other control variables related to features from platform registration processes, which provide further details to the evaluators. A dummy variable is include if a venture has an active website (**Website**) as a proxy of visibility (Koning et al., 2022). I also control for the presence of a video (**Video**), and whether the founding team has submitted a business plan on the project page (**Business Plan**). These attributes offer important information to the crowd about the product or service being developed, as well as plans for expansion in the following years; thus, they contribute to the evaluations (Prokop and Wang, 2021).

Table 2. Summary Statistics

	Obs.	Mean	St. Dev.	Min	Max
<b>Panel A: Project level</b>					
Rating	407	7.282	0.973	3.643	9.857
Investment Interest	407	0.432	0.496	0	1
Female	407	0.265	0.442	0	1
Gender-Identity	359	0.668	0.215	0	1
N. Entrepreneurs	407	1.499	0.759	1	5
Venture Age	405	2.104	8.594	0	170
Incorporated	407	0.622	0.486	0	1
Incubator/Accelerator	405	0.343	0.475	0	1
Prior Funding	407	0.182	0.386	0	1
Foreign Venture	407	0.044	0.206	0	1
Age Team	403	34.961	9.736	19	71
Video	407	0.074	0.262	0	1
Business Plan	407	0.275	0.447	0	1
Website	407	0.828	0.378	0	1
Idea	402	0.025	0.156	0	1
Incubation	402	0.107	0.309	0	1
Development	402	0.269	0.444	0	1
Product Market Fit	402	0.182	0.386	0	1
Launched	402	0.418	0.494	0	1
<b>Panel B: Evaluator level</b>					
N. Evaluators	2,226				
N. Evaluations	4,893	32.230	26.727	1	99
Previous Evaluation	4,893	0.557	0.497	0	1
Female	2,214	0.404	0.491	0	1

Note: This table presents project characteristics at the project (Panel A) and evaluator level (Panel B).



### 3.5 Summary Statistics

Table 2 shows summary statistics for the characteristics at the project level (Panel A) and evaluator level (Panel B). Among the founding teams, 26.5% consist exclusively of women. The distribution of entrepreneurs per venture is skewed, with most early-stage projects ranging from one (63%), two (26%), or three (8%) members. The ventures are relatively young, with an average age of two years. The incorporation rate is high, with 62% of startups registered in the commercial registry. Additionally, 34% of startups have participated in an incubator or accelerator program, and 18% received funding from business angels or venture capitalists at least one year prior to seeking evaluator feedback. Most ventures (95.6%) are based in France. The average age of founding team members is 35 years. Overall, 7% of startups have posted a video on their project page, 27.5% have a business plan, and 82.8% maintain an active website. On the evaluator side, 2,226 distinct evaluators assess an average of 32 projects each. Notably, 40% of evaluators are women, a higher proportion compared to 6% of female venture capitalists and 20% of female angel investors (Brush et al., 2014; Sohl, 2018).

Table 8 presents the correlation matrix, suggesting no excessive correlation between variables used in regression. This is confirmed by the variance inflation factor (VIF) of 1.14, which is below the rule-of-thumb of 10.

## 4 Empirical Strategy

To investigate the gender gap in early-stage project evaluations in terms of rating and investment interest, I estimate the following empirical specification using ordinary least squares (OLS):

$$Y_{ijt} = \beta Female_i + \mathbf{X}_i + \gamma_j + \gamma_t + \varepsilon_{ijt} \quad (1)$$

where  $Y_{ijt}$  stands for the standardized rating or investment interest for project  $i$  assessed by evaluator  $j$  in cohort  $t$ .  $Female_i$  is a dummy variable indicating whether project  $i$  is female-led. The main coefficient of interest is  $\beta$ , which captures differences in early-stage project evaluations. A negative estimated coefficient indicates that evaluators are biased toward female-led startups, which receive lower ratings for their business ideas than male-led startups.  $\mathbf{X}_i$  is a vector of project characteristics that controls for quality, as described in Section 2.4. All control variables were measured in years prior to evaluations to avoid simultaneity bias.  $\gamma_j$  denotes



evaluator fixed effects to control for time-invariant unobserved evaluator characteristics that may confound the identification of  $\beta$ .  $\gamma_t$  denotes cohort fixed effects that control for changes in quality differences across cohorts of startups that join the platform, and  $\varepsilon_{ijt}$  is the error term. To address potential heteroskedasticity and serial correlation, standard errors are clustered at the evaluator level  $j$ , following the suggestion by [Bertrand et al. \(2004\)](#).<sup>8</sup>

To further explore whether the gender identity of the evaluators shares with the gender of the founding teams (i.e., gender homophily) explains the differences in the scores received, I follow [Ahrsjö et al. \(2022\)](#) and estimate the following specification:

$$Y_{ijt} = \beta \text{Gender-Identity}_{ij} + \mathbf{X}_i + \gamma_j + \gamma_t + \varepsilon_{ijt} \quad (2)$$

where the variable Gender-Identity<sub>ij</sub> is an indicator of the extent to which evaluators align with the gender composition of the project founders (i.e., female and male). The indicator ranges from 0 to 1. The coefficient  $\beta$  captures the effect of shared gender identity on project evaluation. A positive estimated coefficient suggests that having more evaluators of the same gender as the founders leads to an increase in the score received.

Despite the baseline specification control for observable characteristics that account for project quality, one potential threat to the interpretation of the gender gap is that the assignment of evaluators to projects is non-random. There are many reasons why a project can be evaluated well or poorly by the crowd. This might depend on the preferences of individual evaluators or the time available to assess a business idea, which I did not observe in the data. For instance, male evaluators may be more likely to discount projects led by female entrepreneurs. I provide evidence that the evaluators are not randomly assigned in Appendix Table 9. Specifically, I regress the gender of founders on a set of project characteristics at the project-evaluator level. I control for cohort fixed effects and cluster standard errors at the evaluator level. There are statistically significant coefficients with an F-test of joint significance of all coefficients that are not equal to zero ( $p < 0.000$ ), suggesting selection of evaluators to projects on observable characteristics.

The non-randomness assignment of evaluators to projects can be partially explained by several characteristics that draw attention to the business idea proposed. For instance, previ-

---

<sup>8</sup>Standard errors are clustered at the evaluator level to account for the possibility that residuals in evaluators' rating decisions may be correlated across projects. In further robustness checks, I verify that the results are robust to not clustering the standard errors and alternative clustering choices.

ous industrial or entrepreneurial experience, market orientation, or access to risky financing could differ between female and male entrepreneurs. These differences can be reflected in the scores and would result in biased OLS estimates. I evaluate this possibility by examining the observed differences between female-led and male-led early-stage startups at the project level. Using information disclosed by the founders on the platform and other information likely to be observed from other sources, I examine several characteristics that might be correlated with the likelihood of subsequent success. Figure 1 shows the mean differences between female- and male-led projects and the confidence intervals of the related t-tests. I find that many characteristics are well balanced. Female-led projects had fewer founders and were less likely to post a video on their profile, a business plan, or a website to present and commercialize their product/service.<sup>9</sup> The other characteristics of female- and male-led businesses are statistically identical and comparable in magnitude, reducing concerns about omitted variable bias in estimating gender differences in project evaluations while adjusting for individual attributes. In robustness checks, I estimate the effect of gender gap using an entropy matching reweighting procedure (Hainmueller, 2012).

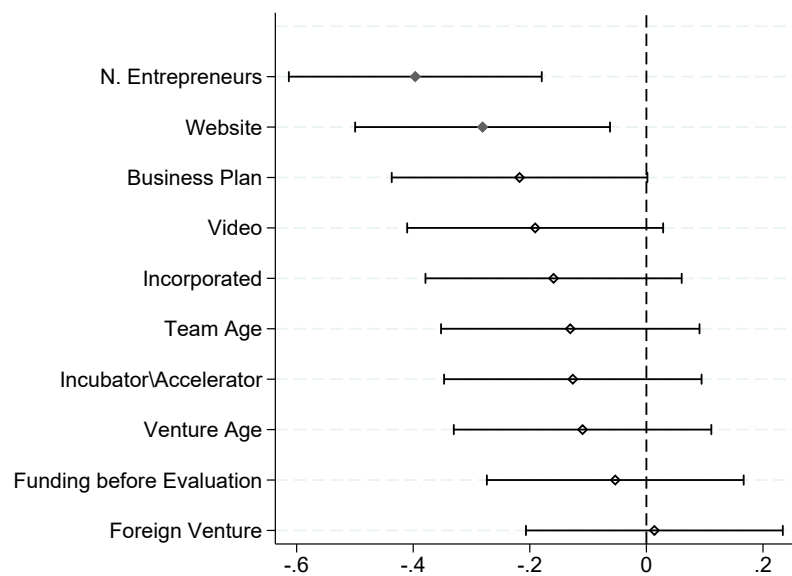


Figure 1. Balancing Test

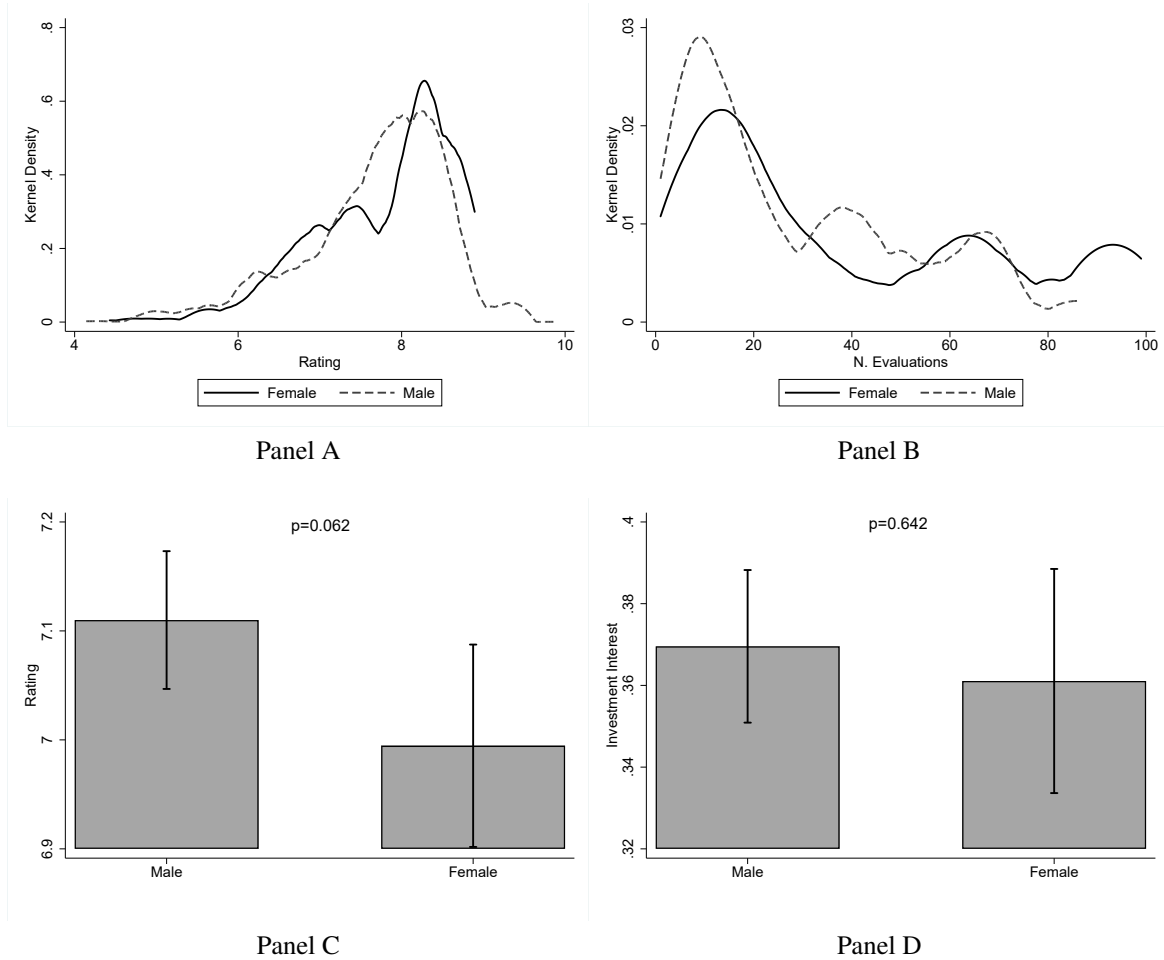


Figure 2. Gender Gap in Evaluations and Investment Interest

## 5 Results

Before proceeding with the analysis of the gender gap in evaluation from the baseline specification, I present descriptive evidence in Fig. 2. This figure compares gender gaps in evaluation scores and the number of evaluations (i.e., reviews) by the crowd in the analysis sample. Reassuringly, distribution of evaluation scores (Panel A) and number of evaluations (Panel B) between teams composed of only women and men follow similar patterns, despite male-led ventures receiving more favorable assessments. In Panel C, I find a small significant gender gap in scores provided by the crowd (-0.115, s.e.= 0.062). In contrast, the bars in Panel D show no significant gender gap in investment interest expressed by the crowd.

In the rest of the paper, I examine the extent to which gender gaps exist in project evaluations in this sample and, whether the composition of the crowd benefits female-led ven-

<sup>9</sup>However, the normalized difference between the two groups is close to the recommended threshold of 0.25 (Imbens and Rubin, 2015), with the exception of the size of the founding team.

tures.

## 5.1 Gender Gap in Project Evaluations

Table 3 shows the results for the effect of the gender of team members on both the standardized z-score and investment interest. Column 1 of Table 3 accounts only for cohort fixed effects and shows a statistically significant negative effect of female-led startups on rating ( $p < 0.1$ ). Evaluators tend to assign scores that are 0.07 SD lower to female-led startups than to male-led startups. Column 2 includes evaluator fixed effects to control for time-invariant unobserved evaluator characteristics. Gender gap in evaluations is similar in magnitude (0.08 SD,  $p < 0.1$ ). These findings indicate that evaluators exhibit bias toward early-stage projects that are solely led by male founders, suggesting that the same ideas receive different scores from the crowd based on the gender of the founders. These results are in line with previous evidence, highlighting gender-based disparities in evaluating entrepreneurial ventures (Brooks et al., 2014; Lee and Huang, 2018; MacNell et al., 2014).

To what extent can gender bias observed in startups be attributed to differences in observable characteristics between women and men? In column 3, I control for project quality attributes to account for potential biases in evaluator assessments. Projects led by women received 0.03 lower SD evaluation scores than those of the other projects. However, the estimated coefficient is statistically insignificant at conventional levels (CI [-0.106; 0.054]). This differs from the unconditional regression, where female-led startups are less likely to receive better evaluation scores. This result suggests that the characteristics of early-stage projects dissipate gender bias against female founders. This indicates the importance of information in mitigating gender stereotypes during crowd evaluations.

In the remaining columns (4-6), I estimate the difference in investment willingness between female-led and those led by men. When controlling only cohort fixed effects, I find a negative effect of gender composition on the probability of evaluators' investment willingness, suggesting that female-led startups are 4 percentage points less likely to receive investment intention than, male-led startups. However, after controlling for evaluator fixed effects and project characteristics, the estimated effects remain negative but turn statistically insignificant at the 10% level. These results suggest that female-led startups receive the same intention to invest in their projects as male-led startups, which confirms previous findings on evaluation scores.

The signs of the estimated coefficients on the observable characteristics included in columns 3 and 6 are consistent with expectations. For instance, venture incorporation, participating in an incubator or accelerator program, and having a business plan have positive and statistically significant effects on the evaluation ratings. In contrast, having a website has a significant negative effect on evaluation scores. For investment interest, only ventures that have been incorporated have a positive effect, while other characteristics are not statistically associated with the probability of investment interest.

Table 3. Effects of Gender on Evaluations

Dep. Var.	Rating			Investment Interest		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.072*	-0.079*	-0.026	-0.049**	-0.013	-0.005
	(0.037)	(0.041)	(0.041)	(0.022)	(0.022)	(0.024)
N. Evaluations	0.011***	0.002**	0.001**	0.003***	0.001*	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)
N. Entrepreneurs			-0.006			-0.002
			(0.019)			(0.010)
Venture Age			-0.0004			0.0001
			(0.001)			(0.001)
Incorporated			0.173***			0.049***
			(0.037)			(0.017)
Incubator/Accelerator			0.116***			0.005
			(0.037)			(0.016)
Prior Funding			0.029			-0.005
			(0.032)			(0.022)
Foreign Venture			0.125			0.026
			(0.093)			(0.046)
Video			0.084			-0.008
			(0.052)			(0.025)
Business Plan			0.130**			0.014
			(0.056)			(0.033)
Website			-0.082**			0.004
			(0.033)			(0.016)
Obs.	4,860	4,860	4,824	4,860	4,860	4,824
R <sup>2</sup>	0.085	0.413	0.428	0.071	0.375	0.376
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Next, I explore whether there is a gender gap in early-stage project evaluations based on the development stage, which is related to the maturity and uncertainty of projects. In Panels A and B of Fig. 3, I estimate the gender gap over the five stages of development from ideas to

Table 4. Effects of Gender on Sub-Scores

Dep. Var.	Success (1)	Useful (2)	Originality (3)	Ambition (4)	Feasibility (5)	Affection (6)	CSR (7)
Female	-0.080* (0.044)	-0.046 (0.048)	0.023 (0.044)	-0.186*** (0.046)	-0.099** (0.039)	-0.011 (0.044)	0.185*** (0.054)
Obs.	4,824	4,824	4,824	4,824	4,824	4,824	4,824
R <sup>2</sup>	0.288	0.222	0.332	0.366	0.396	0.283	0.220
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

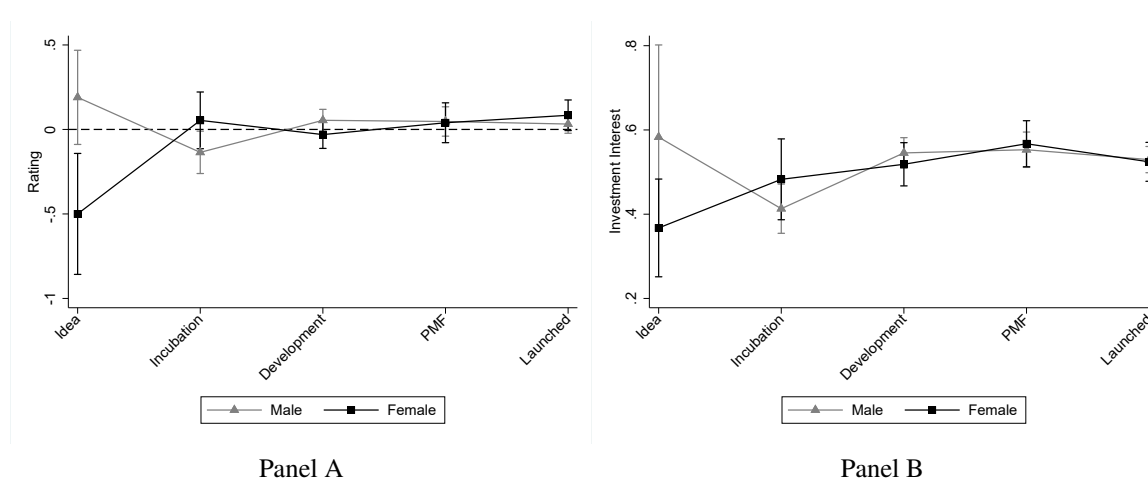


Figure 3. Heterogeneous Effects by Development Stage

products and services launched on the market. Panel A shows that evaluators tend to misbelieve the ideas proposed by women compared to their counterparts. This result suggests that among projects in which uncertainty associated with the business idea is higher because of the absence of concrete signals of development (i.e., business plans or prototypes), the crowd relied on stereotypes and penalized women. In contrast, the evaluation scores received by the projects, at least at the development stage are similar regardless of gender. Similarly, as shown in Panel B of Fig. 3, ideas submitted by women are less likely to receive investment willingness from evaluators, even if the estimated difference is lower.

So far, the findings have shown that projects are evaluated similarly regardless of the gender composition of the founding team. However, the evaluation process is based on several sub-scores that can be assessed differently according to the gender of the team. For instance, evaluators can be harsher toward women on criteria reflecting the operationalization of a business idea and its success than on aspects reflecting other qualities (i.e., economic expectations)

because entrepreneurship is stereotypically viewed as masculine (Eagly and Karau, 2002; Lee and Huang, 2018). At the same time, women-led projects can be less disadvantaged by evaluators on criteria associated with feminine expectations (i.e., social impact). To examine this issue, I estimate whether evaluators assess sub-criteria differently by using a standardized z-score measure.

Table 4 shows the results for the effect of the gender of team members on the six criteria that composed the evaluation, controlling for both cohort, evaluator fixed effects, and observable characteristics. The estimated effects in columns 1 and 4 are negative and statistically significant at conventional levels, implying that female-led projects are viewed as less likely to succeed and to feasibly implement their business ideas than projects founded by men. The evaluation scores received for these two sub-criteria are 0.08 SD and 0.1 SD lower than those of male-led startups (or around 0.63 and 0.75 lower rating, respectively), suggesting a substantial negative gender gap. In contrast, social and environmental concerns included in business ideas developed by women are more likely to receive higher evaluation scores, with a positive gender gap of around 0.19 SD (or around 1.36 higher rating). This latter finding is in line with previous evidence on the diversity of women on boards of directors and their focus on environmental, social, and governance strategy (Purkayastha, 2022), whereas male-led teams tend to focus more heavily on R&D investments (Apesteguia et al., 2012).

Taken together, these findings suggest that there is no significant bias from the crowd for early-stage startups founded by women in terms of both evaluation scores and likelihood of receiving investment interest, conditional on quality-related characteristics. These findings contrast with previous evidence that women are penalized in evaluations compared to men (Snellman and Solal, 2023) and the gender funding gap (Ewens and Townsend, 2020; Guzman and Kacperczyk, 2019). Therefore, leveraging the crowd to assess the underlying potential of early-stage projects can mitigate gender biases or stereotypes, thus suggesting the effectiveness of the wisdom of the crowd in closing the gender gap. However, these results also mask differences in evaluation scores according to criteria. Businesses run by men are assessed as having a greater chance of success and the feasibility of achieving their ideas, whereas female-led ventures outperform in environmental and social challenges.

## 5.2 The effect of evaluators' gender identity on project evaluations

The gender of the evaluators is also an important factor in decision-making and can shape the evaluation scores. Evaluators who share the same gender as the founding team could be more lenient in their perception of project quality. However, evidence on how gender identity affects women's outcomes is mixed. For instance, in venture capital markets dominated by men, male entrepreneurs are more likely to receive external financing because of gender homophily or taste-based discrimination due to investor beliefs (Ewens and Townsend, 2020; Guzman and Kacperczyk, 2019; Snellman and Solal, 2023). However, in academia, Card et al. (2019) find that female-authored papers are similarly assessed by male and female referees. Finally, Bagues et al. (2017) find no evidence that evaluation committees composed only of women favor promoting female candidates.

Table 5. Effects of Gender Identity

Dep. Var.	Rating	Investment Interest	Success	Useful	Originality	Ambition	Feasibility	Affection	CSR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gender Identity	0.015 (0.083)	-0.021 (0.047)	0.024 (0.098)	0.047 (0.095)	-0.021 (0.088)	0.163* (0.084)	0.140* (0.078)	-0.020 (0.097)	-0.188* (0.097)
Obs.	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824
R <sup>2</sup>	0.438	0.376	0.287	0.221	0.332	0.362	0.395	0.283	0.217
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The data allows me to determine whether the gender composition of the evaluators (i.e., homophily) affects score differences. Table 5 shows results from estimating equation 2, with aggregate scores, a binary indicator of whether evaluators are willing to invest, and sub-scores. I include cohort, evaluator fixed effects, and controls for project quality. First, the point estimate in column 1 contrasts with the negative gender gap in Table 3. This finding suggests that the number of evaluators of the same gender as the founding team increase the assessment received, although the estimated coefficient is statistically insignificant (CI [-0.147; 0.179]). The same conclusion emerges for other dependent variables, except for both feasibility and corporate social and environmental sub-scores. I find that evaluators are also more prone to giving better feasibility or lower corporate social and environmental assessments to projects of similar gender (columns 6 and 8, significant at the 5% level).

To further explore heterogeneity related to the gender of the crowd, I explore the role of



gender identity by splitting the sample according to the gender of the founding team in Table 6. I find a positive and significant effect on evaluation scores among female-led startups, whereas the corresponding effect for males is close to zero and statistically insignificant. Female evaluators give better scores for female-led startups. Finally, I find no gender bias in the investment decision.

Taken together, the gender identity of the evaluators had a negligible effect on the project evaluations of the women and men founders, confirming the previous results of no gender differences in crowd evaluations. However, a crowd composed only of women tends to positively favor projects carried out by women, suggesting gender homophily among women evaluators who may apply less stringent criteria when evaluating female-led ventures.

Table 6. Heterogeneous Effects of Gender Identity

Dep. Var.	Score		Interest	
	Female (1)	Male (2)	Female (3)	Male (4)
Gender Identity	0.421** (0.186)	0.039 (0.131)	0.069 (0.105)	-0.052 (0.079)
Obs.	1,636	3,188	1,636	3,188
R <sup>2</sup>	0.534	0.501	0.424	0.385
Year FEs	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

### 5.3 Robustness Checks

The main threat to the empirical strategy is that the OLS estimates could be biased because of the non-random assignment of evaluators to projects. There are three potential concerns with the estimated gender gap in the baseline specification: (i) the observable characteristics available to platform users are not well balanced across female- and male-led projects; (ii) the estimated effects may reflect unobservable characteristics correlated with the gender of the founding team; or (iii) how the dependent variables are measured. Thus, I consider several robustness checks in this section.

As documented in Section 3, a few notable dimensions of female- and male-led projects are not well balanced, such as the number of founders, the presence of an active website, a busi-

ness plan or videos on the platform. Unbalanced characteristics can lead to biased estimates if they are systematically correlated with the dependent variable. To address this concern, I construct a set of weights using an entropy-matching reweighting procedure to make the samples comparable (Hainmueller, 2012).<sup>10</sup> I reweight observations to minimize first-, second-, and third-order moment dimensions for the set of observable characteristics presented in Section 2.4.<sup>11</sup> Appendix Fig. 4 shows the balancing tests of the main characteristics after reweighting, which show identical mean across both groups. Appendix Table 11 shows that the results are qualitatively similar to baseline specifications without perfectly balanced characteristics. The main exception is the statistically significant gender gap in the evaluation using the specification that reweight the female-led projects in the sample to have the same distribution of observables as the male-led projects (0.08 SD,  $p < 0.1$ ).

Another possibility is that the control characteristics of the project quality may not be sufficiently detailed. I further add observable characteristics of quality that include measures of the number of words in the pitch, whether the project is located in an entrepreneurial cluster, whether the team included at least one student and a serial entrepreneur, whether an evaluator made an evaluation prior to the current project, and indicators for the stage of development and industries. In addition, I use the post-double-selection Lasso of Belloni et al. (2014) to select observable characteristics related to project quality rather than making ad hoc decisions about which variables to include. Among the 38 variables, this procedure selected 20 characteristics to control project quality. Appendix Table 12 reports the results, where Panel A includes the full set of controls, and Panel B includes the subset of controls selected from the post-double-selection Lasso. For all dependent variables, I find qualitatively similar results although they are lower in magnitude when controlling for the subset of variables.

I conduct an additional test to assess the potential for bias in omitted variables in the Appendix Table 13. Following Oster (2019), I estimate the bias adjusted coefficients using assumptions about selection bias on unobservable. Assuming a delta of 1, implying that unobservables are equally important for treatment selection as observables, the adjusted coefficient for the standardized crowd evaluation is -0.195. In contrast, assuming a delta of -1 (i.e., un-

---

<sup>10</sup>The entropy matching reweighting procedure presents several advantages over other propensity score approaches. By imposing a larger set of moments (i.e., first, second, and third) of the characteristic distributions, entropy matching achieves a high degree of covariate balance. Moreover, in finite samples, entropy weights performed better than the balance of observable properties.

<sup>11</sup>The matching procedure is performed at the project level to ensure a balance of characteristics. The results are qualitatively comparable for a sample of matched female- and male-led ventures at the project-evaluator level, considering the gender of the evaluators.

observables are more important than observables), the adjusted coefficient is 0.024. Thus, the bounds include zero, which suggests no significant effect as in the baseline specification. In contrast, the bounds for success's criterion are [-0.230; -0.035], which are wider than in Table 4 but exclude zero. Overall, this robustness check provides further evidence that the results are unlikely to be driven by omitted variables.

Finally, I check the robustness of the results with respect to alternative functional forms and samples. In Appendix Table 14, I replace the standardized scores with the nominal ratings. Without or with control variables, the results confirm the conclusions of the baseline specification. In Appendix Table 15, I include mixed-gender majority-male or majority-female teams (106 projects). Conclusions remain unchanged with the alternative sample.

## 6 Discussion and Conclusion

The proportion of women participation in entrepreneurship is lower than that of men, with a lower likelihood of success as entrepreneurs. Most research on the gender gap in entrepreneurship focuses on the funding stage, emphasizing the role of stereotypes in expert decision-making (Ewens and Townsend, 2020; Guzman and Kacperczyk, 2019). However, an unexplored explanation for this gap is the differential evaluation of business ideas between men and women, which may hinder women's efforts to launch new ventures. In this paper, I investigate the gender gap in the evaluation of early-stage projects by leveraging crowd-based assessments. Using novel data from an online platform, I test whether evaluations from a large audience can reduce gender disparities in project assessments.

I find a significant gender gap of 0.08 SD for a sample of ventures at different entrepreneurial stages. In contrast, conditioning on projects' characteristics available to the crowd reduces the estimated gender gap to 0.02 SD, turning insignificant at conventional levels. I further show that women are evaluated less favorably when they propose only one idea, which is a step that does not involve costly expenditure in the development of their project. Finally, I show that projects led by older women are perceived as less viable than their counterparts. At the same time, the composition of the audience and whether they share a higher gender share with the founding teams do not affect evaluation scores or investment interests. These effects are small in magnitude and statistically insignificant. However, female-led projects tend to be positively evaluated when the crowd includes only female evaluators, while male evaluators are not more

favorable toward male-led projects. This finding suggests homophily in the entrepreneurship context, which can enhance female participation in entrepreneurship by narrowing the gender gap.

This paper contributes to several lines of research. First, this paper contributes to the literature on the gender gap in evaluation of early-stage projects within a nonpecuniary context. Previous work examines the assessment of early-stage ventures in the context of resource providers from professional investors, such as VCs or business angels (Brooks et al., 2014; Ewens and Townsend, 2020), pitch competitions (Kanze et al., 2018), and alternative context as crowdfunding campaigns (Bapna and Ganco, 2021; Gafni et al., 2020). This strand of the literature find evidence that female founders are penalized compared to their male counterparts. Gender biases in evaluations are primarily due to stereotypes because of uncertainty and a lack of hard information on the viability of projects (Huang and Pearce, 2015). This paper complements existing research by shedding light on novel evidence that distributed online platforms involving a large audience of nonexpert stakeholders can narrow the gender gap in the evaluation of business ideas at different stages. In addition, observing both the founder and evaluator's gender enables me to explore homophily, suggesting limited evidence of in-group bias in gender. Therefore, crowd assessment could be an effective tool for evaluating early-stage projects and promoting a greater representation of women founders.

Finally, this paper contributes to the growing literature on gender stereotypical beliefs in entrepreneurship. Previous research has demonstrated that female founders experienced less favorable judgment than their male counterparts, leading to financing constraints that limit female participation in entrepreneurship. One noticeable explanation for this negative bias is that female founders are perceived as less competent, irrespective of the quality of their business ideas (Fiske et al., 2002; Yang et al., 2020). By examining evaluators' perceptions across multiple sub-criteria that form the aggregate score, this study highlights heterogeneity in evaluations, specifically in economic potential and social attributes, beyond the final decision (i.e., rating). This paper extends the existing literature by exploring how gender stereotypes influence the decision-making of crowd evaluators in a costless evaluation context. The findings show that male-led projects receive higher scores on criteria aligned with male-associated traits, such as success, ambition, and feasibility, while female-led projects are rated lower on these criteria. Conversely, female-led ventures receive higher evaluations on criteria linked to stereotypically female attributes, such as social and environmental impact. The findings document that the

disadvantages faced by women in entrepreneurship are shaped by gender-stereotyped beliefs. This suggests that evaluating potential based on a single criterion, in which all information has to be integrated into a judgment of a single criterion, can be detrimental.

The results suggest that relying on a diverse audience of nonexperts may be effective for assessing the quality of entrepreneurial projects and improving the representation of women. On the one hand, aggregating a large number of opinions into ratings or other forms may reduce biases compared to a small group of experts and provide relevant information to founders. This can complement other settings, such as blinding founder-related information in idea evaluation without imposing monetary costs (Dahlander et al., 2023b). On the other hand, stakeholders involved may not evaluate ventures in the same manner as professional investors, nor may they have the same motivations or preferences (Boudreau et al., 2016). However, due to uncertainty, even experts experience difficulty screening ideas or proposals and are prone to bias. Future research may explore differences between experts and crowd evaluations and explore potential complementarities to narrow the gender gap.

The paper is not without limitations, opening avenues for future research. First, the exploration of the gender gap in crowd evaluation provides correlation but no causal interpretation of the effects, although several approaches are used to reduce concerns about omitted variable bias. Future research assigning evaluators randomly to female- and male-led entrepreneurial projects could resolve this issue and reinforce the implications for policymakers. Second, the data were obtained from a single online platform in France, which limits the generalizability of the findings. Although the sample has several advantages, such as observing each evaluator's score and gender, enabling me to control for and explore the role of shared identity between founders and evaluators, the findings I present should nevertheless be fairly representative of the population of early-stage entrepreneurs around the world. Future research could explore the extent to which crowd evaluation narrows gender gaps in other institutional settings, such as in the U.S., where gender disparities in entrepreneurship are high. Finally, while the findings document a mechanism for reducing the negative bias against women in idea selection, the data limitations did not allow me to explore whether female- and male-led projects differ at subsequent stages of entrepreneurship and help increase the representation of women. Efforts to collect longitudinal data may help examine the effect of initial screening on long-term outcomes, such as survival rate.

## References

- Abraham M, Botelho T and Dobbin-Lamont G (2024) The (re) production of inequality in evaluations: A unifying framework outlining the drivers of gender and racial differences in evaluative outcomes. *Research in Organizational Behavior* .
- Ahrsjö U, Niknami S, Palme M et al. (2022) Identity in court decision-making. Technical report.
- Aldrich H (1999) *Organizations evolving*. Sage.
- Aldrich HE and Yang T (2013) How do entrepreneurs know what to do? learning and organizing in new ventures. *Journal of Evolutionary Economics* 24(1): 59–82. DOI: 10.1007/s00191-013-0320-x.
- Apestequia J, Azmat G and Iriberrri N (2012) The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science* 58(1): 78–93. DOI:10.1287/mnsc.1110.1348.
- Bagues M, Sylos-Labini M and Zinovyeva N (2017) Does the gender composition of scientific committees matter? *American Economic Review* 107(4): 1207–1238. DOI:10.1257/aer.20151211.
- Bapna S (2019) Complementarity of signals in early-stage equity investment decisions: Evidence from a randomized field experiment. *Management Science* 65(2): 933–952. DOI: 10.1287/mnsc.2017.2833.
- Bapna S and Ganco M (2021) Gender gaps in equity crowdfunding: Evidence from a randomized field experiment. *Management Science* 67(5): 2679–2710. DOI:10.1287/mnsc.2020.3644.
- Bayerl A, Dover Y, Riemer H and Shapira D (2024) Gender rating gap in online reviews. *Nature Human Behaviour* DOI:10.1038/s41562-024-02003-6.
- Belloni A, Chernozhukov V and Hansen C (2014) High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* 28(2): 29–50. DOI: 10.1257/jep.28.2.29.
- Bernstein S and Korteweg K Arthur and Laws (2017) Attracting early-stage investors: Evidence from a randomized field experiment. *The Journal of Finance* 72(2): 509–538. DOI:10.1111/jofi.12470.
- Bertrand M, Duflo E and Mullainathan S (2004) How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119(1): 249–275. DOI:10.

1162/003355304772839588.

- Bordalo P, Coffman K, Gennaioli N and Shleifer A (2016) Stereotypes\*. *The Quarterly Journal of Economics* 131(4): 1753–1794. DOI:10.1093/qje/qjw029.
- Boudreau KJ, Guinan EC, Lakhani KR and Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science* 62(10): 2765–2783. DOI:10.1287/mnsc.2015.2285.
- Brooks AW, Huang L, Kearney SW and Murray FE (2014) Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences* 111(12): 4427–4431. DOI:10.1073/pnas.1321202111.
- Brush C, Greene PG, Balachandra L and Davis AE (2014) Diana report. women entrepreneurs 2014: Bridging the gender gap in venture capital. arthur m. *Blank Center for Entrepreneurship, Babson College* .
- Brush CG, Duhaime IM, Gartner WB, Stewart A, Katz JA, Hitt MA, Alvarez SA, Meyer GD and Venkataraman S (2003) Doctoral education in the field of entrepreneurship. *Journal of management* 29(3): 309–331.
- Card D, DellaVigna S, Funk P and Iriberry N (2019) Are referees and editors in economics gender neutral?\*. *The Quarterly Journal of Economics* 135(1): 269–327. DOI:10.1093/qje/qjz035.
- Carter S and Rosa P (1998) The financing of male– and female–owned businesses. *Entrepreneurship & Regional Development* 10(3): 225–242. DOI:10.1080/08985629800000013.
- Cook C, Diamond R, Hall JV, List JA and Oyer P (2020) The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *The Review of Economic Studies* 88(5): 2210–2238. DOI:10.1093/restud/rdaa081.
- Correll SJ and Benard S (2006) *Biased estimators? Comparing status and statistical theories of gender discrimination*. Emerald (MCB UP ), pp. 89–116. DOI:10.1016/s0882-6145(06)23004-2.
- Cumming D, Meoli M and Vismara S (2019) Does equity crowdfunding democratize entrepreneurial finance? *Small Business Economics* 56(2): 533–552. DOI:10.1007/s11187-019-00188-z.
- Dahlander L, Beretta M, Thomas A, Kazemi S, Fenger MH and Frederiksen L (2023a) Weeding out or picking winners in open innovation? factors driving multi-stage crowd selection on

- lego ideas. *Research Policy* 52(10): 104875. DOI:10.1016/j.respol.2023.104875.
- Dahlander L, Thomas A, Wallin MW and Ångström RC (2023b) Blinded by the person? experimental evidence from idea evaluation. *Strategic Management Journal* 44(10): 2443–2459. DOI:10.1002/smj.3501.
- Dardenne B, Dumont M and Bollier T (2007) Insidious dangers of benevolent sexism: Consequences for women’s performance. *Journal of Personality and Social Psychology* 93(5): 764–779. DOI:10.1037/0022-3514.93.5.764.
- Davidsson P, Grégoire DA and Lex M (2021) Venture idea assessment (via): Development of a needed concept, measure, and research agenda. *Journal of Business Venturing* 36(5): 106130. DOI:10.1016/j.jbusvent.2021.106130.
- Dutt N and Kaplan S (2018) Acceleration as mitigation: When processes can address gender bias in entrepreneurship. *Academy of Management Proceedings* 2018(1): 16160. DOI: 10.5465/ambpp.2018.16160abstract.
- Eagly AH and Karau SJ (2002) Role congruity theory of prejudice toward female leaders. *Psychological Review* 109(3): 573–598. DOI:10.1037/0033-295x.109.3.573.
- Ewens M and Townsend RR (2020) Are early stage investors biased against women? *Journal of Financial Economics* 135(3): 653–677. DOI:10.1016/j.jfineco.2019.07.002.
- Fiske ST, Cuddy AJC, Glick P and Xu J (2002) A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82(6): 878–902. DOI:10.1037/0022-3514.82.6.878.
- Franzoni C and Sauermann H (2014) Crowd science: The organization of scientific research in open collaborative projects. *Research Policy* 43(1): 1–20. DOI:10.1016/j.respol.2013.07.005.
- Gafni H, Marom D, Robb A and Sade O (2020) Gender dynamics in crowdfunding (kickstarter): Evidence on entrepreneurs, backers, and taste-based discrimination\*. *Review of Finance* 25(2): 235–274. DOI:10.1093/rof/rfaa041.
- Galton F (1907) Vox populi. *Nature* 75(1949): 450–451. DOI:10.1038/075450a0.
- Goldin C and Rouse C (2000) Orchestrating impartiality: The impact of “blind” auditions on female musicians. *American Economic Review* 90(4): 715–741. DOI:10.1257/aer.90.4.715.
- Greenberg J and Mollick E (2017) Activist choice homophily and the crowdfunding of female founders. *Administrative Science Quarterly* 62(2): 341–374. DOI:10.1177/0001839216678847.



- Gross DP (2020) Creativity under fire: The effects of competition on creative production. *The Review of Economics and Statistics* 102(3): 583–599. DOI:10.1162/rest\_a\_00831.
- Gupta VK, Goktan AB and Gunay G (2014) Gender differences in evaluation of new business opportunity: A stereotype threat perspective. *Journal of Business Venturing* 29(2): 273–288. DOI:10.1016/j.jbusvent.2013.02.002.
- Guzman J and Kacperczyk AO (2019) Gender gap in entrepreneurship. *Research Policy* 48(7): 1666–1680. DOI:10.1016/j.respol.2019.03.012.
- Hainmueller J (2012) Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20(1): 25–46. DOI: 10.1093/pan/mpr025.
- Hall RE and Woodward SE (2010) The burden of the nondiversifiable risk of entrepreneurship. *American Economic Review* 100(3): 1163–1194. DOI:10.1257/aer.100.3.1163.
- Haynie JM, Shepherd DA and McMullen JS (2009) An opportunity for me? the role of resources in opportunity evaluation decisions. *Journal of Management Studies* 46(3): 337–361. DOI:10.1111/j.1467-6486.2009.00824.x.
- Hegde D and Sampat B (2015) Can private money buy public science? disease group lobbying and federal funding for biomedical research. *Management Science* 61(10): 2281–2298. DOI: 10.1287/mnsc.2014.2107.
- Hilton JL and von Hippel W (1996) Stereotypes. *Annual Review of Psychology* 47(1): 237–271. DOI:10.1146/annurev.psych.47.1.237.
- Hoornaert S, Ballings M, Malthouse EC and Van den Poel D (2017) Identifying new product ideas: Waiting for the wisdom of the crowd or screening ideas in real time. *Journal of Product Innovation Management* 34(5): 580–597. DOI:10.1111/jpim.12396.
- Huang L and Pearce JL (2015) Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions. *Administrative Science Quarterly* 60(4): 634–670. DOI:10.1177/0001839215597270.
- Imbens GW and Rubin DB (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. ISBN 9781139025751. DOI: 10.1017/cbo9781139025751.
- Johnson MA, Stevenson RM and Letwin CR (2018) A woman’s place is in the... startup! crowdfunder judgments, implicit bias, and the stereotype content model. *Journal of Business Venturing* 33(6): 813–831. DOI:10.1016/j.jbusvent.2018.04.003.

- Kanze D, Conley MA, Okimoto TG, Phillips DJ and Merluzzi J (2020) Evidence that investors penalize female founders for lack of industry fit. *Science Advances* 6(48). DOI:10.1126/sciadv.abd7664.
- Kanze D, Huang L, Conley MA and Higgins ET (2018) We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal* 61(2): 586–614. DOI:10.5465/amj.2016.1215.
- Kolev J, Fuentes-Medel Y and Murray F (2019) *Is Blinded Review Enough? How Gendered Outcomes Arise Even Under Anonymous Evaluation*. DOI:10.3386/w25759.
- Koning R, Hasan S and Chatterji A (2022) Experimentation and start-up performance: Evidence from a/b testing. *Management Science* 68(9): 6434–6453. DOI:10.1287/mnsc.2021.4209.
- Lakhani KR (2016) The antidote to hippos: Crowd voting. *Harvard Business Review* 2.
- Lee M and Huang L (2018) Gender bias, social impact framing, and evaluation of entrepreneurial ventures. *Organization Science* 29(1): 1–16. DOI:10.1287/orsc.2017.1172.
- Li D (2017) Expertise versus bias in evaluation: Evidence from the nih. *American Economic Journal: Applied Economics* 9(2): 60–92. DOI:10.1257/app.20150421.
- Li D and Agha L (2015) Big names or big ideas: Do peer-review panels select the best science proposals? *Science* 348(6233): 434–438. DOI:10.1126/science.aaa0185.
- Luo H, Macher J and Wahlen M (2021) Judgment aggregation in creative production: Evidence from the movie industry. *Management Science* 67(10): 6358–6377. DOI:10.1287/mnsc.2020.3815.
- MacNell L, Driscoll A and Hunt AN (2014) What’s in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education* 40(4): 291–303. DOI: 10.1007/s10755-014-9313-4.
- Magnusson PR, Wästlund E and Netz J (2014) Exploring users’ appropriateness as a proxy for experts when screening new product/service ideas. *Journal of Product Innovation Management* 33(1): 4–18. DOI:10.1111/jpim.12251.
- Mitchell JR and Shepherd DA (2010) To thine own self be true: Images of self, images of opportunity, and entrepreneurial action. *Journal of Business Venturing* 25(1): 138–154. DOI:10.1016/j.jbusvent.2008.08.001.
- Mollick E and Nanda R (2016) Wisdom or madness? comparing crowds with expert evaluation in funding the arts. *Management Science* 62(6): 1533–1553. DOI:10.1287/mnsc.2015.2207.

- Müller-Trede J, Choshen-Hillel S, Barneron M and Yaniv I (2018) The wisdom of crowds in matters of taste. *Management Science* 64(4): 1779–1803. DOI:10.1287/mnsc.2016.2660.
- Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2): 187–204.
- Pitchbook (2022) All in: Female founders in the us vc ecosystem .
- Prokop J and Wang D (2021) Is there a gender gap in equity-based crowdfunding? *Small Business Economics* 59(3): 1219–1244. DOI:10.1007/s11187-021-00574-6.
- Proserpio D, Troncoso I and Valsesia F (2021) Does gender matter? the effect of management responses on reviewing behavior. *Marketing Science* DOI:10.1287/mksc.2021.1313.
- Purkayastha A (2022) How can women board members bring focus on the governance of esg strategy? *California Management Review Insights* .
- Quignon A (2023) Crowd-based feedback and early-stage entrepreneurial performance: Evidence from a digital platform. *Research Policy* 52(7): 104792. DOI:10.1016/j.respol.2023.104792.
- Reimers I and Waldfogel J (2021) Digitization and pre-purchase information: The causal and welfare impacts of reviews and crowd ratings. *American Economic Review* 111(6): 1944–1971. DOI:10.1257/aer.20200153.
- Reskin BF (2003) Including mechanisms in our models of ascriptive inequality: 2002 presidential address. *American Sociological Review* 68(1): 1. DOI:10.2307/3088900.
- Riedl C, Blohm I, Leimeister JM and Krcmar H (2010) Rating scales for collective intelligence in innovation communities: Why quick and easy decision making does not get it right. In: *Proceedings of Thirty First International Conference on Information Systems*.
- Riedl C, Blohm I, Leimeister JM and Krcmar H (2013) The effect of rating scales on decision quality and user attitudes in online innovation communities. *International Journal of Electronic Commerce* 17(3): 7–36. DOI:10.2753/jec1086-4415170301.
- Santamaría L and Mihaljević H (2018) Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* 4: e156. DOI:10.7717/peerj-cs.156.
- Sauermann H, Vohland K, Antoniou V, Balázs B, Göbel C, Karatzas K, Mooney P, Perelló J, Ponti M, Samson R and Winter S (2020) Citizen science and sustainability transitions. *Research Policy* 49(5): 103978. DOI:10.1016/j.respol.2020.103978.
- Scott EL and Shu P (2017) Gender gap in high-growth ventures: Evidence from a university venture mentoring program. *American Economic Review* 107(5): 308–311. DOI:10.1257/

aer.p20171009.

- Scott EL, Shu P and Lubynsky RM (2020) Entrepreneurial uncertainty and expert evaluation: An empirical analysis. *Management Science* 66(3): 1278–1299. DOI:10.1287/mnsc.2018.3244.
- Shafi K (2019) Investors' evaluation criteria in equity crowdfunding. *Small Business Economics* 56(1): 3–37. DOI:10.1007/s11187-019-00227-9.
- Simcoe TS and Waguespack DM (2011) Status, quality, and attention: What's in a (missing) name? *Management Science* 57(2): 274–290. DOI:10.1287/mnsc.1100.1270.
- Smith KG, Baum JR and Locke EA (2001) A multidimensional model of venture growth. *Academy of Management Journal* 44(2): 292–303. DOI:10.2307/3069456.
- Snellman K and Solal I (2023) Does investor gender matter for the success of female entrepreneurs? gender homophily and the stigma of incompetence in entrepreneurial finance. *Organization Science* 34(2): 680–699. DOI:10.1287/orsc.2022.1594.
- Sohl J (2018) The angel investor market in 2017: Angels remain bullish for seed and start-up investing .
- Surowiecki J (2005) *The Wisdom of Crowds*. Anchor. ISBN 9780385721707.
- Thébaud S (2015) Business as plan b: Institutional foundations of gender inequality in entrepreneurship across 24 industrialized countries. *Administrative Science Quarterly* 60(4): 671–711. DOI:10.1177/0001839215591627.
- Wright NL, Koning R and Khanna T (2023) Judging foreign startups. *Strategic Management Journal* 44(9): 2195–2225. DOI:10.1002/smj.3490.
- Wu Z and Chua JH (2011) Second-order gender effects: The case of u.s. small business borrowing cost. *Entrepreneurship Theory and Practice* 36(3): 443–463. DOI:10.1111/j.1540-6520.2012.00503.x.
- Yang S, Kher R and Newbert SL (2020) What signals matter for social startups? it depends: The influence of gender role congruity on social impact accelerator selection decisions. *Journal of Business Venturing* 35(2): 105932. DOI:10.1016/j.jbusvent.2019.03.001.
- Yu S (2020) How do accelerators impact the performance of high-technology ventures? *Management Science* 66(2): 530–552. DOI:10.1287/mnsc.2018.3256.
- Zhang J (2009) The advantage of experienced start-up founders in venture capital acquisition: evidence from serial entrepreneurs. *Small Business Economics* 36(2): 187–208. DOI:10.1007/s11187-009-9216-4.

## 7 Supplementary Material

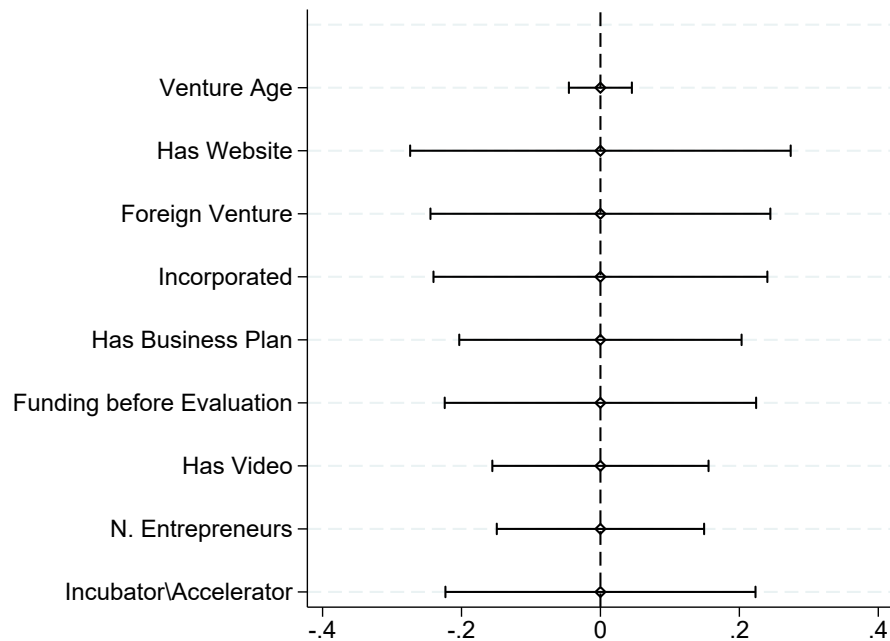


Figure 4. Balancing Tests after Entropy Matching

Table 7. Variable Definition and Sources

Variable	Description	Source
<b>Dependent Variables</b>		
Rating	Standardized evaluator's aggregate score	Wirate
Investment Interest	Binary indicator for evaluator's expression of investment interest	Wirate
<b>Independent Variables</b>		
Female	Binary indicator equal to one if the founding team is composed only by female and zero otherwise	Wirate
Gender-Identity	Indicator for how many of the evaluators share same gender with the founding team	Author's calculation
<b>Control Variables</b>		
N. Entrepreneurs	Size of the founding team	Wirate
Venture Age	Venture age measured at the launch date on the platform	Wirate and Register of commerce
Incorporated	Binary indicator equal to one if the venture is incorporated and zero otherwise	Register of commerce and myfrenchstartup
Incubator/Accelerator	Binary indicator equal to one if the venture has been incubated or accelerated and zero otherwise	Dealroom/myfrenchstartup
Prior Funding	Binary indicator equal to one if the venture raised funds	Dealroom/myfrenchstartup
Foreign Venture	Binary indicator equal to one if the venture is located outside France	Wirate
Age Team	Average age of the founding team	Register of commerce/LinkedIn
Video	Binary indicator equal to one if the venture has a video and zero otherwise	Wirate
Business Plan	Binary indicator if the venture has a business plan and zero otherwise	Wirate
Website	Binary indicator if the venture has an active website and zero otherwise	Dealroom/myfrenchstartup/LinkedIn
Previous Evaluations	Binary indicator equal to one if an evaluator has previously assessed a project and zero otherwise	Wirate

Table 8. Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Rating	1.000														
(2) Investment Interest	0.556***	1.000													
(3) Female	0.034**	-0.030***	1.000												
(4) Gender-Identity	-0.054***	-0.037**	-0.406***	1.000											
(5) N. Entrepreneurs	0.000	0.019	-0.325***	0.112***	1.000										
(6) Venture Age	-0.045***	-0.027*	-0.057***	0.047***	-0.057***	1.000									
(7) Incorporated	-0.003	0.014	-0.216***	0.163***	0.011	-0.011	1.000								
(8) Incubator/Accelerator	0.019	0.003	-0.179***	-0.018	0.088***	-0.039***	0.224***	1.000							
(9) Financing before Round	-0.013	-0.033**	-0.062***	0.059***	0.047***	-0.000	0.101***	0.049***	1.000						
(10) Foreign Venture	0.069***	0.052***	0.083***	-0.123***	0.063***	0.194***	-0.304***	-0.193***	-0.051***	1.000					
(11) Video	0.024*	0.000	-0.109***	0.016	0.194***	-0.014	-0.072***	-0.097***	0.071***	0.197***	1.000				
(12) Business Plan	-0.065***	-0.037***	-0.138***	0.036**	-0.015	0.111***	0.101***	-0.070***	0.023	-0.067***	0.102***	1.000			
(13) Website	0.001	0.004	-0.229***	0.151***	0.137***	0.069***	0.385***	0.028*	0.096***	-0.033**	0.069***	0.021	1.000		
(14) Age Team	-0.084***	-0.063***	-0.061***	0.077***	-0.259***	0.144***	0.110***	-0.142***	-0.009	-0.096***	0.036**	0.210***	0.069***	1.000	
(15) Experienced Evaluator	-0.423***	-0.353***	0.009	0.078***	-0.038***	0.031**	0.035**	0.009	0.040***	-0.073***	-0.005	-0.001	0.040***	0.093***	1.000

Table 9. Random Assignment of Evaluators

	Female
N. Entrepreneurs	-0.196*** (0.008)
Venture Age	-0.003*** (0.000)
Incorporated	-0.096*** (0.016)
Incubator/Accelerator	-0.101*** (0.014)
Financing before Evaluation	-0.014 (0.015)
Foreign Venture	0.113*** (0.023)
Website	-0.083*** (0.019)
Video	-0.188*** (0.026)
Business Plan	-0.126*** (0.012)
Obs.	4,824
Joint F-test	[0.000]

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 10. Balancing Venture Characteristics

	Female	Male	p-value	Norm. Diff.
N. Entrepreneurs	1.278	1.579	0.000***	0.423
Venture Age	1.411	2.352	0.119	0.132
Incorporated	0.565	0.642	0.164	0.158
Incubator/Accelerator	0.299	0.359	0.254	0.128
Financing before Evaluation	0.167	0.187	0.628	0.054
Foreign Venture	0.046	0.043	0.905	-0.014
Team Age	34.024	35.295	0.249	0.136
Video	0.037	0.087	0.042**	0.208
Business Plan	0.204	0.301	0.040**	0.225
Website	0.750	0.856	0.024**	0.269

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 11. Effects of Gender on Evaluations: Entropy Matching

Dep. Var.	Rating (1)	Investment Interest (2)	Success (3)	Useful (4)	Originality (5)	Ambition (6)	Feasibility (7)	Affection (8)	CSR (9)
Female	-0.046 (0.038)	-0.004 (0.024)	-0.099** (0.047)	-0.056 (0.052)	0.020 (0.043)	-0.194*** (0.042)	-0.123*** (0.037)	-0.013 (0.042)	0.138*** (0.052)
Obs.	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824
R <sup>2</sup>	0.461	0.405	0.312	0.238	0.355	0.416	0.440	0.312	0.246
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 12. Effects of Gender on Evaluations

Dep. Var.	Rating (1)	Investment Interest (2)	Success (3)	Useful (4)	Originality (5)	Ambition (6)	Feasibility (7)	Affection (8)	CSR (9)
<b>Panel A: Additional Characteristics</b>									
Female	0.010 (0.043)	0.003 (0.026)	-0.084* (0.047)	0.045 (0.047)	-0.018 (0.050)	-0.133*** (0.044)	-0.092** (0.044)	0.022 (0.045)	0.248*** (0.059)
<b>Panel B: Lasso Selected Characteristics</b>									
Female	-0.029 (0.043)	-0.002 (0.026)	-0.080* (0.047)	0.001 (0.047)	-0.053 (0.051)	-0.188*** (0.048)	-0.085** (0.043)	-0.005 (0.047)	0.180*** (0.052)
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 13. Effects of Gender on Evaluations: Oster Test

Dep. Var.	Rating (1)	Investment Interest (2)	Success (3)	Useful (4)	Originality (5)	Ambition (6)	Feasibility (7)	Affection (8)	CSR (9)
<b>Panel A: Baseline Effect</b>									
Female	-0.026 (0.041)	-0.005 (0.024)	-0.080* (0.044)	-0.046 (0.048)	0.023 (0.044)	-0.186*** (0.045)	-0.099** (0.039)	-0.011 (0.044)	0.185*** (0.054)
<b>Panel B: Oster bounds</b>									
Oster (2019) bounds $\beta$	[-0.195; 0.024]	[-0.023; 0.055]	[-0.230; -0.035]	[-0.176; -0.007]	[-0.122; 0.065]	[-0.408; -0.118]	[-0.207; -0.068]	[-0.084; 0.011]	[0.012; 0.234]
$\delta$ Selection ratio for $\beta=0$	0.840	0.072	1.343	1.462	-1.943	2.315	0.975	0.202	1.501
Obs.	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,284
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 14. Effects of Gender on Evaluations: Nominal Rating

Dep. Var.	Rating	Success	Originality	Useful	Ambition	Feasibility	Affection	CSR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A: No Controls</b>								
Female	-0.137* (0.072)	-0.272*** (0.087)	-0.207** (0.094)	-0.031 (0.087)	-0.421*** (0.086)	-0.319*** (0.080)	-0.110 (0.094)	0.399*** (0.121)
<b>Panel B: Baseline Characteristics</b>								
Female	-0.045 (0.071)	-0.158* (0.087)	-0.089 (0.094)	0.049 (0.096)	-0.350*** (0.085)	-0.205** (0.080)	-0.024 (0.095)	0.459*** (0.134)
Mean of the Dep. Var.	7.734	7.833	8.044	7.713	7.849	7.499	7.823	7.377
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 15. Effects of Gender on Evaluations: Mix-Gender Teams

Dep. Var.	Rating	Investment Interest	Success	Originality	Useful	Ambition	Feasibility	Affection	CSR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.013 (0.038)	-0.003 (0.021)	-0.070* (0.039)	0.017 (0.040)	-0.015 (0.043)	-0.186*** (0.045)	-0.076** (0.035)	-0.007 (0.040)	0.177*** (0.051)
Obs.	5,806	5,806	5,806	5,806	5,806	5,806	5,806	5,806	5,806
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$