Do Employees Work Less for Female Leaders?
A Multi-Method Study of Entrepreneurial Firms

Olenka Kacperczyk
okacperczyk@london.edu
London Business School

Peter Younkin
younkin@uoregon.edu
Lundquist College of Business, University of Oregon

Vera Rocha
vr.si@cbs.dk
Copenhagen Business School

Abstract

We propose that female-founded ventures receive a lower amount of employee labor for equal pay because employees are more likely to decline requests for additional labor by female founders. First, using longitudinal matched employer-employee data covering all founders of new ventures with personnel in Portugal between 2002 and 2012, we confirm that full-time employees contribute fewer regular hours and less overtime work to female-founded firms. Second, using a series of online experiments, we show that this variation in employee labor across female and male-founded firms is partly motivated by a difference in the employee's expectations of work demands. Specifically, employees perceive female founders' requests for additional labor to be unfair and more difficult than expected, and both of these perceptions explain the lower amount of employee labor supplied in female-founded ventures. Overall, our findings uncover a novel mechanism that helps explain the existence of a gender gap in entrepreneurship beyond the entry stage.

Keywords: Entrepreneurship, Gender, Leadership, Staffing and HR

Acknowledgements: The author team is grateful for the constructive comments of three anonymous reviewers and the senior editor, and for all the feedback received on earlier versions of this paper presented at INSEAD, ESMT, Copenhagen Business School, and Santa Clara University.

1. INTRODUCTION

Despite the public, political, and academic interest in reducing gender inequality in leadership positions, female leaders continue to face persistent disadvantages relative to their male counterparts (England, 1992; Hillman, Cannella and Harris, 2002; Naumovska, Wernicke, and Zajac, 2020). Beyond the known difficulties in advancing to leadership roles, recent work has also highlighted the range of individual and organizational challenges that follow once women reach these positions (Kacperczyk, Kang and Paik, 2021; Solal and Snellman, 2019). There has been hope that entrepreneurship might offer a means to moderate these gaps, given that female founders can access leadership positions more easily in their own firms (Yang, Kacperczyk and Naldi, 2021) and may be inclined to create better environments for future female leaders (Rocha and van Praag, 2020). A long line of work on gender and entrepreneurship, however, has shown that women confront greater scrutiny when founding organizations and, for reasons that remain unclear, many of these organizations struggle to perform in the long run (Kim, Aldrich, and Keister, 2006; Ruef, Aldrich, and Carter, 2003; Yang and Aldrich, 2014). As a result, the potential for entrepreneurship to offer women an alternate path to leadership roles remains largely unrealized.

Research on gender inequality in entrepreneurship has examined the obstacles that female founders face at entry, especially when they seek early-stage investment (Castellaneta, Conti and Kacperczyk, 2020; Guzman and Kacperczyk, 2019; Thébaud, 2010). Yet theories of gender inequality at entry are unlikely to fully explain why female founders continue to face adversity post-entry. For example, the most common mechanism responsible for gender disparities at founding – the "evaluative bias" of key resource holders, who rely on gender as a proxy for female founders' quality or commitment (Balachandra et al., 2019; Bird and Brush, 2002; Buttner and Rosen, 1989; Thébaud, 2015b; Thébaud and Sharkey, 2016) – seems inconsistent with persistent hardship post-founding for two reasons. First, if resource providers (e.g., investors, banks, and crowds) apply greater scrutiny to female founders, then the average quality and commitment of those receiving support would exceed that of male founders. Second, gender is most often used as a proxy for quality and commitment when uncertainty is high (Correll, 2001; Correll, Benard and Paik, 2007), but this uncertainty is expected to decline as founders offer more reliable quality signals in later stages of new venture development. Thus, any disparities between female and male founders should also decline over time, which is inconsistent with the patterns observed. The need to identify alternative sources of gender disadvantage in entrepreneurship post-entry, or once women occupy leadership positions in new ventures, is what motivates our study.

To resolve these inconsistencies between the theories about female disadvantage at founding and the empirical patterns observed post-founding, we integrate a growing line of research on the role of employees for startup growth and success (e.g., Agarwal, 2019; Agarwal, Braguinsky and Ohyama, 2019; DeSantola and Gulati, 2017) with sociological theories of the gendering of managerial and leadership roles (Cullen and Perez-Truglia, 2021; Eagly, Karau and Makhijani, 1995; Eagly and Karau, 2002) to identify a novel source of bias, which arises when employees assess the volume of labor they will supply to a given firm depending on the founder's gender. Specifically, we propose that, relative to male-founded startups, ventures founded by women will elicit a lower amount of employee labor for equal pay, leading to what we term an *employee labor imbalance*. Extending recent work on how gender influences employees' perception of managers (Ranganathan and Shivaram, 2020) or leaders (Beaman et al., 2009) and research on the key influence of founders in a new venture (Burton and Beckman, 2008; Campero and Kacperczyk, 2019), we further attribute this labor imbalance to employees' biased inferences about reduced workload when the founder is a woman. In particular, given the widespread cultural beliefs that women prioritize work less than men (Acker, 1990; Jacobs and Gerson, 2004; Reid, 2015), we suggest that start-up employees will formulate different expectations for the work demands in each environment. Specifically, we expect that employees will associate female founders with lower work demands and will thus commit less work time for the same pay, partly because they will find requests for extra work less acceptable and more difficult than anticipated when the founder is a woman instead of a man.

To test our theory, we take a two-step approach. First, we use employer-employee matched data from Portugal for the period 2002-2012 to assess whether an employee's total amount of labor provided to a startup (at a given pay) varies with its founder's gender. We complement these analyses with two experiments, in which we hired temporary workers through an online platform to complete work for a fictional startup, manipulating the apparent gender of the founders. This approach provides direct insight into the behavior of workers that are commonly employed by startups (Cardon, 2003) and parallels recent studies that have used online workers to understand the motivation for employee behavior more broadly, net of any unobserved supply-side variations (Burbano, 2016; Leung, 2014). Furthermore, the use of experiments allows us to examine the mechanisms driving the employee labor imbalance in female-led ventures and confirms the robustness of this effect.

Overall, our study makes multiple contributions. First, we contribute to female entrepreneurship research by going beyond its predominant focus on women's hardship experienced at entry stages (e.g., Coleman and Robb, 2009; Guzman and Kacperczyk, 2019; Thébaud, 2010). We document a new source of female founder disadvantage that occurs post-entry: an imbalance in employee labor supply. Second, we extend recent studies on how gender shapes subordinate-supervisor interactions by documenting that employees may, under some conditions, negatively stereotype female supervisors in ways that make them less compliant with requests for labor (Cullen and Perez-Truglia, 2021; Eagly et al., 1995, Eagly and Karau, 2002; Ranganathan and Shivaram, 2020). Finally, we contribute to the long line of inquiry on gender gaps in career outcomes (Cha and Weeden, 2014; Cotter et al., 1997; England, 1992; Huffman and Velasco, 1997; Reskin, 2000). Whereas prior research has mostly focused on discrimination of female subordinates by their supervisors, we instead show that subordinates also discriminate against their female superiors. In this respect, our study shows that discrimination may be more prevalent than previously thought, originating not only in upper-level positions but also in lower-ranked positions.

2. THEORY

2.1. Past Research

A long tradition of research has documented that women in leadership positions such as managerial or executive roles face persistent disadvantages relative to men (e.g., England, 1992; Reskin, 2000; Heilman, 1980, 1984; Quadlin, 2018). Though entrepreneurship is often promoted as a means for excluded individuals to attain positions of leadership, research increasingly suggests that the hardship that female leaders face is present even among start-up founders. For example, female founders tend to underperform and have less influence and power within founding teams relative to their male peers (Kim et al., 2006; Ruef et al., 2003; Yang and Aldrich, 2014). Despite this empirical pattern, existing theories do not fully explain why such gap persists after women have managed to surmount the initial entry barriers and launched a new venture. Current theories of female entrepreneurship apply predominantly to the founding stage and do not easily extend to post-entry stages. The most frequent explanation for the gender gap in entrepreneurship rates is the bias of financial resource providers (e.g., venture capitalists, loan officers, angel investors, or banks) when they evaluate founders at the critical entry stage (Bigelow et al., 2014; Brooks et al., 2014; Buttner and Rosen, 1989; Thébaud, 2015b; Thébaud and Sharkey, 2016). Women are generally seen as less credible, less competent, and less committed founders (Buttner and Rosen, 1988; Thébaud, 2015b; Balachandra et al., 2019; Bird and Brush, 2002) because, in the absence of other indicators, resource holders use gender to infer the underlying quality of the new firm or the founder's ability and willingness to commit to it (Benard and Correll, 2010; Castilla, 2008; Castilla and Benard, 2010; Correll et al., 2007). Hence, prospective female founders face greater scrutiny than male founders and such scrutiny reflects the evaluators' attribution of uncertainty to founder's gender.

While pertinent to founding stages, this evaluative bias is unlikely to drive the persistent gender disadvantage post-entry. If female founders are subject to higher standards and greater scrutiny at entry, they must then demonstrate a higher average quality than male founders to acquire the resources necessary to launch. Moreover, as a new venture matures or grows, the uncertainty about its quality should decline, reducing the importance of demographic cues such as founder's gender in evaluators' deliberations of quality or commitment (Heilman, 1984; Heilman et al., 1988). Yet most female founders remain disadvantaged post-entry (Kim et al., 2006; Ruef et al. 2003; Yang and Aldrich, 2014) and this is incongruent with theories of evaluative bias, which would imply equal or even higher (not lower) performance among those female founders who pass the initial scrutiny of resource providers.

To resolve this contradiction between extant theories and empirical findings, we turn our attention to challenges arising post-entry and identify new sources of female disadvantage during the growth stage, once the entry barriers have been circumvented. An emergent stream of research establishes early employees and their labor as a key resource for startup growth (e.g., Agarwal, 2019; Agarwal et al., 2019; DeSantola and Gulati, 2017; Hietaniemi et al., 2021; Honoré and Ganco, 2020), with employee motivation and effort being especially critical to the day-to-day operations of new ventures (Sauermann, 2018). We build on this research and extend recent work on how a manager's gender affects employee behavior (Ranganathan and Shivaram, 2020, Abraham and Burbano, 2021) to develop a framework that theorizes differences in employee labor across female and male-founded startups.

2.2. The Value of Employee Labor for Startup Advantage and the Role of Founders

There is ample evidence that entrepreneurial success hinges on the founders' ability to motivate their employees, including eliciting their effort to successfully develop a product, establish routines, and scale (e.g., Agarwal, 2019; DeSantola and Gulati, 2017). Yet start-ups often lack the resources of established firms to incentivize employee effort (Burton, Dahl, and Sorenson, 2018; Sauermann, 2018). Consequently, young, fledgling ventures often face unique challenges to elicit employee labor and maximize employee output (Stinchcombe, 1965; Aldrich and Ruef, 1965). This trade-off between eliciting employee labor and minimizing costs becomes more difficult as startups experiment with their strategy and structure (Sine, Mitsuhashi and Kirsch, 2006), which may require extra effort and commitment from their employees to an extent that is hard to specify ex-ante, at the time of hiring. In face of these constraints, being able to flexibly elicit labor when needed without increasing costs is critical to startup survival. Indeed, prior work shows that startup employees often work long hours (Livingston, Beth and Timothy, 2008; Lussier, 1995), sometimes well beyond what was specified in their initial contracts (Litwin and Phan, 2013), and without commensurate immediate pay (Burton et al., 2018; Sorenson et al., 2021). In sum, variation in a founder's ability to elicit the preferred employee behaviors (i.e., motivation, stability, commitment), including enticing workers to supply greater labor without adjusting their monetary rewards, constitutes a significant competitive advantage.

Traditionally, organizations use a mixture of intrinsic and extrinsic incentives to motivate employees to increase their investment in the firm. However the same financial constraints and absence of formal structures (Stinchcombe, 1965; Tolbert et al., 2011) that amplify the value of employee labor also limit the incentives available in startup firms. For example, immediate financial or career payoffs (e.g., bonuses or promotions) are largely uncertain in startups (Kacperczyk and Marx, 2015; Sorenson et al., 2021), as failure or insufficient growth are common outcomes. Therefore, rather than financial incentives, it is often the non-pecuniary benefits of working for a startup that enhance employee motivation in these settings (Roach and Sauermann, 2015; Sauermann, 2018). Furthermore and as a result, startup employees themselves are more likely than employees of large organizations to determine their effort or commitment by making inferences based on the founder's attributes. Indeed, recent research shows that founders play a critical role in influencing employee behavior (Agarwal, 2019) – without well-defined structures and incentives, founders are central in establishing the routines and values of a new firm (Agarwal et al., 2019; Campero and Kacperczyk, 2019; Roach and Sauermann, 2015). This founder imprinting effect extends to choices about how work is regulated, which behaviors are rewarded and how (Baron and Hannan, 2002). In addition, the founders' identities and backgrounds can further shape insiders' and outsiders' assessments of the firm (Hallen, 2008; Hsu, 2007), including employees' perceptions (Baron and Hannan, 2002; Campero and Kacperczyk, 2019). Employees will thus be liable to base their expectations of a firm's work environment, at least partly, on their impression of the founder.

In sum, because startups often lack established structures and incentive schemes, employees' perceptions, assessments, and behaviors can vary depending on who the founder is. A critical question is thus whether and how the founder's gender will influence employees' inferences about the workplace.

2.3. Employee Labor and Founder Gender

Prior work on the gendering of leadership roles suggests that, on average, employees are attuned to the gender of an organization's founder or leader (e.g., Cohen and Huffman, 2007; Reskin, 2000). Although there is no systematic consensus on how differently men and women manage organizations (see Chapman, 1975; Eagly et al., 1995; Eagly and Karau, 2002), it has been found that individuals *anticipate* differences based on the leader's gender (e.g., Abraham and Burbano, 2021). For example, widely-held cultural beliefs imply that leadership roles require masculine traits, which women are believed to lack by being perceived as more agreeable, dependable, and compliant (Heilman, 1980, 1984; Quadlin, 2018), and thus worse fits for upper-level positions compared to men (Eagly and Karau, 2002; Kacperczyk, Kang, and Paik, 2021; Solal and Snellman, 2019). Importantly, in work contexts, even if female managers outperform their male counterparts, employees often expect female managers to perform low-status work but penalize male managers who do the same, suggesting that an employee's notion of what managing entails varies with the manager's gender (Ranganathan and Shivaram, 2020).

These persistent associations between a manager's gender and an employee's expectations will be accentuated in startup firms. Many attributes of a startup founder, from their experience to their personality, influence employees' expectations, and a founder's gender may be a particularly prominent attribute used in this type of subjective appraisals. Key evaluators such as investors (Huang, Land and Pearce, 2015), banks (Irwin and Scott, 2010), crowds (Greenberg and Mollick, 2017), and employees (Campero and Kacperczyk, 2019) have been found to adjust their perception of a firm based on the founder's gender. For example, using data from an online labor market, Campero and Kacperczyk (2019) find that job applicants consider founders' demographic traits in deciding whether to apply for a startup job. Important to the question at hand, if a founder's gender is salient to different audiences and used as a signal of their ability to launch and grow their venture (Buttner and Rosen, 1988; Thébaud, 2010), it might shape the perceptions of employees, namely their anticipation of the firm's work demands.

Sociological research documents a longstanding relationship between gender and perceptions of work demands (Cha, 2010; Cha and Weeden, 2014; Hochschild and Machung, 2012), as employees tend to use gender cues to predict workplace obligations. For example, the "ideal worker" theory (Acker, 1990; Jacobs and Gerson, 2004) posits that women are often perceived as more attentive to family obligations (Bianchi et al., 2000; Correll et al., 2017) and thus as less likely to work beyond their contract (Cha and Weeden, 2014). Extending this argument to startups, the gendered notion of an ideal worker provides one rationale for relevant evaluators, including employees, to form expectations of lower work demands when the firm is founded by a woman. Female founders will likely elicit a commonly-held cultural belief that lower workload or effort is expected, given that women are generally perceived as prioritizing work-life balance over high-devotion labor and firm goals. These broad associations between gender and work demands will be amplified by factors specific to entrepreneurship – such as the common perceptions that (at least some) women start new ventures precisely to accommodate disproportionate work–life demands, including childrearing and household chores (e.g., Brett and Stroh, 2003; Kacperczyk and Younkin, 2021; Thébaud, 2015a). Though women may sometimes become founders in the hope of developing more flexible work schedules, resolving work-family conflicts (Lombard, 2001), or reducing the cost of childcare (e.g., Connelly, 1992; Thébaud, 2015a), these general tendencies become established beliefs even if women's founding motives may differ substantially. For example, using an audit study to probe employers' perceptions of ex-entrepreneurs, Kacperczyk and Younkin (2021) found that prospective employers inferred the accommodation of childcare duties to be the main motive behind women's choice to become founders. By the same logic, employees may hold similar expectations that female-founded ventures are primarily motivated by childcare responsibilities and that they are oriented to advance lifestyle goals (e.g., work–life balance) rather than growth outcomes (Gorman and Kmec, 2009, Hochschild and Machung, 2012). Hence, the presence of a female founder may function as a signal that the firm has lower work demands than an equivalent male-founded firm.

Importantly, these expectations about lower work demands when the founder is a woman will negatively influence the employees' labor supply because requests for additional labor will be perceived as inconsistent with commonly-held norms. Indeed, a large literature suggests that people modify their behaviors, including the effort they are willing to provide at work, to conform to their understanding of local norms (Cialdini and Goldstein, 2004). Critically, if a request appears to be aligned with commonly-held norms, including the beliefs and assumptions about work demands, such request is more likely to be accepted. By contrast, requests inconsistent with broader norms are likely to be rejected. For example, employees are more likely to engage in environmental and pro-social behaviors when they find such acts consistent with workplace norms (Kim et al., 2017). Similarly, employees accept levels of workplace risk disproportionate to any financial reward when they perceive risky behaviors as part of their implicit agreement (Desmond, 2008). Extending this logic to female-led firms, employees may deny requests for additional labor because such requests will be perceived as a violation of their prior assumptions about work demands. More specifically, when expectations about workplace demands are violated, employees may develop negative perceptions of the work itself and thus decline requests for extra labor at higher rates than equivalent employees in firms conforming to pre-established assumptions. This will result in an employee labor imbalance – whereby employees are less willing to supply additional labor – placing female founders at a disadvantage relative to male founders.

In terms of the mechanisms, several negative perceptions may emerge when female founders elicit additional effort from their employees. But perceptions of unfairness and greater work difficulty relative to expectations are especially likely to prevail, justifying why employees deny those requests and supply less labor to female-led startups. Regarding the former explanation, psychology research shows that violations of expectations can trigger judgements that certain behaviors are unfair or lack sufficient justification (e.g., Nowak, Page and Sigmund, 2000) and should therefore be punished (Fehr and Fischbacher, 2004; Rudert and Greifeneder, 2016; Tuscherer et al., 2015). As a result, by violating widely-held beliefs that working for a female-led company is less demanding, female founders' requests for additional work may be perceived as unfair and unjustified. For example, if employees hold pre-conceived notions that female leaders tend to prioritize work-life balance over firm goals and therefore should elicit lower volumes of labor than male leaders, any violation of this belief will trigger the perception of unfairness, leading employees to deny requests for additional work at higher rates when they come from female than from male founders.

Second, to the extent that a founder's gender is associated with pre-established expectations of lower work demands, requests for additional labor may trigger the perception that the work is more difficult than anticipated. Importantly, this perception of greater difficulty relative to expectations will reduce the likelihood of accepting additional labor requests, given that employees' responses critically depend not only on their perceptions of the workplace environment but also of the task or job itself (Abeler et al., 2011; Campbell, 1988). Prior research suggests, for example, that the perceived labor demands partly depend on the resources required to complete the job (Liu and Li, 2012) as well as the *a priori* determinability of the job requirements from the worker's point of view (Byström and Järvelin, 1995). In this sense, a request for more labor can signal a mismatch between the expected input and the actual input required at the job (O'Donnell and Johnson, 2001). From this perspective, jobs that require fewer resources (e.g., effort, time) than anticipated are perceived as easier, while jobs that exceed the anticipated resource threshold are conventionally described as more demanding (Hendy, Liao, and Milgram, 1997). Similarly, employees view tasks that fall outside rather than within the boundaries of an implicit work agreement as more burdensome or even inappropriate (Semmer et al., 2015), and they anticipate that more effort is needed to successfully complete such tasks (Horvath, Herleman, and McKie, 2006). Other research further shows that, if employees consider tasks inappropriate—because their assignment violates the expected norms about work demands or the fit for a particular job or profession—this can lead to counterproductive work behaviors, reducing rather than increasing employee effort (Semmer et al., 2010). At the extreme, turnover rates may increase when employees are assigned tasks they find inappropriate or against their expectations (Eatough et al., 2016). Further, employees' assessments of work demands influence their willingness to engage such that the more difficult the uncompensated request, the less likely an employee may be to comply (Charness, Gneezy and Henderson, 2018). Hence, since a founder's gender can be used as an indication of the workload required, female founders will also influence employees' perceptions about the job's difficulty. An increase in labor demand from a female founder will violate employees' expectations of workload and trigger the perception that working for a female-led startup is more difficult or more demanding than anticipated, thus leading to higher rejection rates and an *employee labor imbalance* relative to male-led startups.

In conclusion, we propose that, because no contract perfectly describes the responsibilities and expectations of any work position, employees will attempt to fill these information gaps by drawing inferences about what is expected of them in terms of labor supply. This will generally work to the disadvantage of female founders because employees are liable to associate female founders with lower work demands and will thus commit less work time for equal pay. More formally, we predict:

> *H1: Startup employees will be more likely to withhold labor (for equal pay) when working for female than for male-founded firms.*

3. EMPIRICAL DESIGN

3.1. Multi-Method Research Design: Observational Data and Two Experiments

The relationship between a founder's gender and employee labor at a given pay is subject to multiple confounding factors. For example, differences in the labor supplied to female and male founders could be spurious if female-founded firms provide employees with weaker financial incentives. Similarly, our correlations could be spurious if employees with lower human capital are less likely to commit their time but more likely to work for female founders. Yet differences in employee and firm quality are often unobserved to researchers and difficult to measure with precision. To alleviate these critical concerns, we adopt a multi-method research design. We begin with large-scale observational data to determine whether there is an employee labor imbalance in female-led startups relative to male-led startups. These data allow us to test our hypothesis in a representative, country-level, setting. We complement these analyses with online experiments that benefit from the random assignment of employees to founders of different gender. Besides reiterating the validity of H1 and identifying the causal link between founder's gender and employee labor, these experiments are central to our understanding of the underlying mechanisms.

3.2. Study 1: Observational Data

3.2.1. Data and Sample

We first use *Quadros the Pessoal* (QP), an employer–employee linked dataset maintained by the Portuguese Ministry of Employment to identify all private firms founded between 2002 and 2012 in Portugal. Following other studies (Dahl and Klepper, 2015; Sauermann, 2018), we focused on all new firms employing at least one wage earner between 18 and 60 years old. Our initial sample included 231,041 firms not older than 5 years. As we focus on startups with personnel, we excluded any corporate ventures, firms for which no founder could be identified, firms with no employees besides the founder(s), and firms with missing information.[1] Our final sample covers 58,832 firms and their size distribution is

---

[1] Similar to prior studies, we implemented the following step-wise deletion of corporate ventures. First, we dropped 2,443 firms with multiple establishments, unlikely to be startups. Second, following other studies using QP data (e.g., Branstetter et al., 2014), we identified founders based on individuals' employment status, i.e. whether they are paid employees or (one of) the employer(s) of the focal firm. We removed 128,794 firms for which no founder could

available in Appendix Table 1A. Finally, we merged this sample with the QP files at the employee-level to prepare a linked employer–employee data set containing 842,738 observations. We dropped 260,321 observations corresponding to either firm founders, given our focus on employees, or missing information at the employee-level. Our final sample contained 582,417 observations pertaining to 243,269 employees.

3.2.2. Dependent Variable

To measure *Employee Labor*, we use rich information on the hours worked by each employee at the respective firm because, for a given pay, employees presumably contribute greater volumes of labor when they work longer. For each individual employed by the time of the annual QP data collection, we extracted both the number of regular hours and the number of extra hours worked at the focal firm per month. Because the number of hours may simply vary with the job contract (part-time or full-time), we limit our analyses to full-time employees.[2] Also, since many employees report zero *Extra Hours*, we assess patterns in this variable conditional on having worked at least one extra hour in the reference month. Zero values for this variable might indicate that the employee was not offered the choice of working extra time or that s/he did but refused. Some zeros in this variable may thus not reflect a supply-side response, which is our main focus. For robustness, we estimate a zero-inflated model (reported in Appendix Table 4A). Finally, we assess differences by founder gender in the duration (in hours) of the regular (weekly) work schedule, which the employee and the employer agree upon when signing the job contract. Higher values indicate that an employee agreed to a longer weekly work schedule.

3.2.3. Independent Variables and Controls

Our main independent variable is a binary variable equal to "1" for *Female Founders*, and "0" for male founders. Theoretically, we do not focus on mixed-gender teams as gender diversity may trigger

---

be identified. We further excluded 14,503 firms for which we could only observe manager(s) but not founder(s). Finally, we excluded firms without personnel and firms with missing values. These steps left us with 58,832 firms.

[2] Since part-time employees may differ in unobservable characteristics that can affect their willingness to supply labor, we focus on full-time employees, who are more homogenous and less subject to omitted variable bias. We provide employee-level descriptive statistics in Table 2. Appendix Table 2A describes employees in female-led firms. For robustness, we re-estimate our main models including part-time employees (see Appendix Table 6A).

conceptually different processes (Ruef et al., 2003; Yang and Aldrich, 2014), leading employees to either discount or reward such firms. However, we isolate those cases by controlling for mixed-gender teams.

We control for several employee characteristics that may correlate both with employee labor and founder's gender. We account for employee's gender, age, immigration status, education level (less than high school, high school, or university education), years in wage employment and years in entrepreneurship (i.e. as employer of a firm with personnel). We also include monthly wages (in log) and rank in the firm because an employee's pay and hierarchy likely influence their effort. The latter is based on the classification of occupations, ranging from managerial roles and high-skill jobs to low-skill jobs (e.g., roles in operations and more manual jobs). At the founder level, we consider their age, experience as an employee and as an entrepreneur. For startups with multiple founders, we compute the mean value of these variables. We further account for founders' education and for foreign-born founders. In addition to all these covariates, our models control for firm-level characteristics, such as firm size (measured as log number of employees), firm age (ranging from 1 to 5), and firm sales (in log of Euros). Finally, we account for idiosyncratic variation in the environment by including regional (county-level), industry (2-digit NACE), and year dummy variables in all models.

3.2.4. Estimation Method

Our unit of analysis is the employee-firm-year. Given the hierarchical structure of the data, we estimate multi-level mixed effects models to account for two kinds of effects: a) fixed effects, i.e., standard regression coefficients describing the population and b) random effects in the form of intercepts that can account for unobserved heterogeneity at various levels of analysis.[3] Because heterogeneous employees are nested within heterogeneous firms, we add firm-level and employee-level intercepts to control for heterogeneity in organizational contexts and acknowledge that firms (and their founders) may shape each employee's behavior in a unique way depending on the correlation between their unobserved characteristics. Formally, our main specification is represented as follows:

[3] Multi-level mixed-effects models account for unobserved heterogeneity at the employee and the firm levels but with less restrictive identification assumptions than the Random Effects estimator (Abowd, Kramarz and Woodcock, 2008).

$$Hours_{ift} = \beta_0 + \beta_1 FF_{if} + X_{it}\alpha + Z_{ft}\delta + \gamma_y + \mu_j + \lambda_r + \zeta_{if}^{(2)} + \zeta_f^{(3)} + \varepsilon_{if} \quad (1)$$

where FF_{if} is equal to "1" if the firm is only led by female founders, "0" otherwise; the vectors X_{it} and Z_{ft} represent respectively employee and firm/founder characteristics, some of which may vary over time; γ_y, μ_j, and λ_c are year, industry (2-digit), and county fixed effects; $\zeta_{if}^{(2)}$ and $\zeta_f^{(3)}$are the employer-by-firm and firm-level random effects with zero mean and variances $\psi^{(2)}$ and $\psi^{(3)}$, respectively; and ε_{if} is the level 1 error term. In consistency checks, we estimate count models given that *Hours* are measured in positive integer values. Finally, we also test the consistency of our results using linear models with employee-fixed effects to further account for employee unobserved characteristics that may correlate with founder gender and the number of hours worked (e.g., ability, ambition, or preferences that remain stable during the period analyzed). All these models cluster the standard errors at the firm-level.

3.3. Study 2: Experimental Data

We use Study 1 to offer population-level evidence of an employee labor imbalance across female- and male-founded startups. However, this study cannot distinguish between supply-side and demand-side mechanisms, i.e., whether female founders simply differ from male founders in the amount of work they demand, or whether the employees themselves offer less labor to female than to male founders. As we try to identify whether employees are less willing to contribute more hours (for the same pay) to female founders, a key empirical challenge lies in accounting for unobserved variation and reverse causality. For example, correlations between employee labor and founder's gender might simply reflect differences between female- and male-founded startups and their employees. Our results could also be spurious if less committed workers would sort into female-led ventures. Without random assignment of employees across startups, we cannot provide causal evidence of our hypothesis. Hence, we conducted an online experiment (which we refer to as Study 2) in which we hired individuals to complete an image coding task resembling a type of repetitive work that is necessary and often outsourced. This allowed us to randomize whether the workers believed these tasks were being done for a team of male or female founders.

We used Amazon's Mechanical Turk (mTurk), an online labor marketplace, to test how an employee's willingness to supply labor for a given pay varies with founders' gender. MTurk is suitable for our experiment for several reasons. First, this online setting has been widely used by organizations to hire workers (i.e. contractors, freelancers, and temporary workers) for precisely this type of repetitive, time-consuming, but necessary labor (Gartside et al., 2013). Indeed, recent studies suggest that startups are particularly likely to use online labor given their acute resource constraints (e.g., Kässi and Lehdonvirta, 2016). This makes mTurk a suitable platform for studying how workers respond to employer differences (e.g., Burbano, 2016; Leung, 2014). Second, mTurk makes it possible to randomly assign workers across firms and to manipulate founders' gender by altering the information provided to each worker regarding their employer. Third, beyond its implications for startups, online workers have become more prominent over time, encouraging management and strategy scholars to investigate online labor markets (Burbano, 2016; Burtch, Carnahan and Greenwood, 2018; Leung, 2014; Leung, 2017).[4]

3.3.1. Subject Recruitment

In December of 2016, we recruited 500 online workers to complete tasks for a startup firm in exchange for pay. To solicit these workers, we used the following script: "*We're the founders of a new tech startup building a photo database and we need some help with image coding*." The work promised $1 payment and estimated the total time required at 8–10 minutes.[5] We only accepted U.S.-based respondents with 500+ prior HITS and a 95%+ approval rating. Online settings impose well-known limitations, including the loss of experimental control, but they also have a notable advantage of allowing researchers to tap the broader population than the one typically recruited to experiments in university labs, which increases external validity (Crump et al., 2013). Nevertheless, to ensure that our results were not driven by low quality participants (Aguinis, Villamor and Ramani, 2020), we repeated our main analyses after excluding participants in the top or bottom 5 percent of photo counts. Our results remained

[4] MTurk workers are found to be demographically representative of the U.S. working population (Berinsky et al., 2012) and their behavior is similar to that of other workers (Horton et al., 2011; Paolacci et al., 2010).
[5] All respondents received payment, irrespective of the quality of their work. The experiment was conducted in compliance with the IRB board at the second author's University (approval #458-0417).

consistent. Finally, we took three steps to increase the perceived authenticity of the request. First, we asked for help with a simple image interpretation, a common job on mTurk.[6] Second, we created a new mTurk requester account linked to a static webpage of a forthcoming startup instead of using our academic accounts. Upon selecting our post, workers were taken to an external website with the startup logo and information about the founders. We also tracked comments on the message boards for mTurk workers to monitor potential contamination and to ensure that respondents believed the work was for a new firm. Third, because pre-screening out respondents might have contaminated the study (suggesting that it was a study and not actual work for a startup), we allowed respondents to participate multiple times but included only their first response in our analyses. Respondents who failed to complete the task, who indicated any problem (e.g., the picture loaded slowly or was blurry), who failed the attention check (described below), and who shared an IP address with a prior respondent were excluded from the analysis (about 9 percent of all respondents). A total of 457 respondents were then used in the analysis. These respondents self-reported as male (47%), white (70%), under 35 (61%), and as having worked at a startup at some point (14%). A full list of the questions asked and their order is available in Appendix B.

3.3.2. Experimental Design

To test our hypothesis – i.e., the existence of an employee labor imbalance across male- and female-led ventures – we adopt a between-subjects design that compares employee labor supply in three different conditions: *Control, Male Founders,* and *Female Founders*. For each condition, respondents received the same set of instructions upon accepting the assignment (see Figure 1). For example, in the *Control* condition, respondents were told "*We recently founded a photography-based startup and need help coding these photographs for our database*." Respondents were then asked to count the persons in a scenic photograph and to enter the number into a field in our survey. This task was repeated for 12 successive photos. The task was designed to be repetitive and mundane, identical in design to many prior "real-effort" experiments used to identify how different primes affect a worker's willingness to complete

[6] Image interpretation, audio transcription, and survey completion are the most frequent tasks posted on mTurk.

or exceed a task (for a review see DellaVigna and Pope, 2016). While the respondents are working on a single task and are thus not comparable to permanent full-time employees, researchers have routinely used this "real effort" approach to understand the underlying dynamics that govern employee-employer relations (e.g., Carpenter, Matthews, Schirm, 2010; Tonin and Vlassopoulos, 2014). In our case, the work is designed to approximate the experience of performing a typical office task, while allowing us to test how founder's gender influences a given worker's response to task-specific requests.

******* Figure 1 about here*******

The main treatment was the manipulation of founders' gender. In the *Control* condition, the prompt was identical to Figure 1 but neither the founders nor the startup were named. We used this condition to establish a baseline behavior for subjects in an unprimed state. In the *Male Founders* treatment, the prompt included a final line stating "Thank You" followed by "Matthew & Joe, Co-Founders of Photolytics." In the *Female Founders* treatment, only the name of the founders was changed to "Amanda & Chloe".[7] After completing the 12 photos, all subjects received a prompt thanking them for their work and asking whether they would be willing ("Yes"/"No") to analyze some extra photos for no additional payment. The request was identical across all three conditions and only varied in the final line, with the *Control* providing no name for the founders and the *Male/Female Founders* conditions listing the names described above. Those willing to continue to work were shown up to five extra photos and could stop at any time. When the respondent indicated that they had completed the task, they were presented with an attention check ("*Do you remember the names of the founders*?"), a set of questions about their

[7] The use of gender-associated first names allows us to test the effect of founder gender while introducing the fewest possible confounding variables. To do so, we had to ensure that the selected names signalled a particular gender but no differences in presumed ethnicity or socioeconomic status (SES). In a pre-test, we solicited 400 respondents via mTurk to "answer a survey of your impression of different names, colleges, and words" and paid $0.20. Consistent with research on appropriate audit-method techniques (Gaddis, 2017), we showed the respondents a random set of four (out of 40 total) names and asked them to indicate whether each person was male/female, white/black, and what level of education they would presume the person's mother attained. We selected four names that strongly indicated a single gender and did not vary in their perceived ethnicity or SES (Appendix Table 10A). In addition, we chose a photo coding task, as photography was not perceived as an either male or female activity.

impressions of the startup, and a final set of demographic questions. These were not required for payment and roughly 20 percent of the respondents failed to answer at least one of the questions.[8]

4. RESULTS

4.1. Testing Hypothesis 1

4.1.1. Observational Evidence

We start by analyzing whether employees working for female and male founders report any different behavior in the raw data used in Study 1. Table 1 presents initial statistics for the three different measures of employee labor by founder gender. In line with our hypothesis, employees in female-founded startups work fewer hours (both regular and extraordinary hours) compared to employees in male-led startups. About 61 (27) percent of the employees in our sample work for male (female) founders and 11 percent work in firms founded by mixed-gender teams (Table 2). Roughly 41 percent are female employees, but this share tends to be greater in female-founded startups (Appendix Table 2A). These patterns indicate that women are less likely to start firms and, once they do, their workforce differs. The average founder in our sample is 38 years old, has almost 12 years of work experience and 5 years of experience as an entrepreneur. Appendix Table 3A provides a correlation matrix for the key variables used in Study 1.

******* Tables 1 and 2 about here *******

Table 3 reports the coefficients from employee-level regression models assessing variation in different measures of *Hours Worked* across female- and male-founded ventures. Models 1–3 are multi-level mixed models and models 4-6 are count data models with standard errors clustered at the firm-level. We find that full-time employees in female-led startups work, on average, fewer regular hours per month than employees in male-led startups (model 1). Model 2 reports a more sizeable difference between employees of female and male founders – conditional on working overtime, the former group works about 1.4 fewer extra hours a month than the latter. Further, as shown in model 3, working for a female-led startup is negatively associated with the regular number of work hours per week. While the gender gap in

[8] A post-hoc analysis showed that the minimum sample required to find an effect with 0.8 power was 228 responses, well below the final number of observations used in our analyses (i.e. 457).

regular hours is relatively small (below 1%), which may be explained by the relatively low employee discretion regarding their regular schedule, the difference in extra hours is more pronounced and amounts to a 7% lower supply of extra labor relative to the average number of hours observed among employees who work overtime. These patterns generally hold when using count models (models 4-6) and when we include employee fixed effects (Table 5A) and use a two-way clustering for the standard errors.[9] Yet given the significant intra-class correlations found in our data, the multi-level models reported in Table 3 are our preferred specification, as they account for the hierarchical clustering present in our data.

******* Table 3 about here *******

Whereas these analyses reveal stark discrepancies in employee labor in female and male-founded startups, a fair concern relates to systematic differences between female and male founders and their employees that may correlate with labor volumes. To mitigate this concern, we re-estimated our models on a matched sample of employees working for male and female founders based on the large set of controls included in Table 3. Appendix Table 7A reports the estimated difference in hours worked by founder gender using Propensity Score Matching (PSM). The results remain virtually unchanged.[10,11]

Overall, Study 1 reveals a significant variation in the hours worked, whereby employees work fewer regular and extra hours in female-founded firms, in line with H1.[12] Although our observational data do not allow us to identify why the variation in labor emerges, nor imply causality, they uncover patterns consistent with the notion that employees may work less for female than for male founders, as theorized.

---

[9] The founder gender gap in employee labor is also generally consistent when we estimate the model with the full sample of employees and control for the type of contract (Appendix Table 6A).

[10] Appendix Table 8A reports the covariate balance after PSM and confirms that treated and control groups became comparable after matching. In alternative, we used Coarsened Exact Matching to exactly match employees in male- and female-led firms in industry (2-digit), region (NUTs 3), firm size quartiles and year. The results remain similar despite the smaller matched sample.

[11] A related concern may be that our findings only apply to life-style ventures that are unlikely to grow. We re-estimate our baseline models for high-growth ventures, i.e. startups with below-average number of employees and above-average sales. Appendix Table 9A shows similar results with this sample.

[12] In extra analyses (available upon request), we did not find any differences by employee's gender.

4.1.2. Experimental Evidence

We next assess whether the patterns found in Study 1 hold in an experimental setting. In Study 2, we assess whether the treatment conditions (*Female* versus *Male Founders*) change the respondent's willingness to work longer (for equal pay). We measured this willingness in two ways. First, the *Likelihood of Extra Labor* is a binary variable coded "1" if the individual agreed to code extra images for the same pay, "0" otherwise. Second, *Extra Work Volume* is the number of photos the respondent coded without any additional pay. Table 4 reports the covariates' balance across different conditions: *Control* (unnamed founders) versus *Male Founders* (Panel A) and *Control* versus *Female Founders* (Panel B).

******* Table 4 about here*******

Table 5 displays the results of regressions predicting both outcomes. Because some respondents did not answer all the demographic questions, we first present a model for all respondents without controlling for respondent characteristics and then report two models that include these variables. Starting from the logit models predicting the *Likelihood of Extra Work*, models 1 and 3 use the *Control* condition as a benchmark and estimate the effect of each prime (*Male* and *Female Founders*) relative to unnamed founders. Models 2 and 4 are restricted to named founders and estimate the difference between *Female Founders* and *Male Founders*. Controlling for respondent characteristics (model 3), we find that subjects were more willing to offer extra labor to male (p<0.001) and female (p<0.001) founders than to unnamed founders. However, respondents were less likely to offer extra labor to a female than a male founder (Chi-Squared=8.85; p=0.003). This finding is consistent across all models. In practical terms, 50% of the respondents faced with the *Male Founders* treatment agree to work extra time for no additional pay, while only 40% of those faced with the *Female Founders* condition do so (for comparison, less than 30% of the respondents in the *Control* condition offered extra work). Appendix Figures 1A and 1B show the predicted probabilities of providing extra work with and without controls, respectively.

In models 5–8, we test the differences in the *Extra Work Volume* completed. As respondents could decide to stop providing additional work at any point, we estimated a count (Poisson) model predicting the total number of extra photos coded (for the same pay). In model 7, we find again that any indication of founder identity increased the volume of extra work, as respondents coded more photos for *Male* (p<0.001) and *Female Founders* (p<0.001) than for unidentified founders, controlling for respondent characteristics. Moreover, the difference between the two coefficients is statistically significant (Chi-Squared=7; p=0.008), with female founders again exhibiting a disadvantage. Model 8 confirms that respondents coded fewer additional photos for female founders than for male founders. The aggregate effect is substantial, as the average respondent coded 1.9 (1.6) extra photos for male (female) founders. This gap amounts to a 2 percent advantage in productivity for male founders. These results are consistent with our hypothesis that female founders are less able to elicit extra effort (for the same pay) from their employees than male founders, which translates into a significant productivity disadvantage.

*******Insert Table 5 about here*******

4.2. Testing Underlying Mechanisms

With our experiment (Study 2), it is possible to rule in one explanation for the pattern observed in the Portuguese data (Study 1), i.e. startup employees are more likely to penalize female founders by withholding time and supplying less labor. To clarify what motivated this behavior, we conducted a follow-up experiment (presented below as Study 3), in which we investigated whether differences in the respondent's perception of work demands mediated our findings (partially or fully).

4.2.1. Study 3: Follow-up Experiment

In September 2017, we solicited 1,200 additional respondents using the same recruitment script, pay, and exclusion criteria as in Study 2. As before, we removed from the sample anyone who had participated in the prior round, who failed the manipulation check, who finished either too quickly or too slowly, and who reported any problem with their experience.[13] A total of 982 new respondents were included in our analyses. Most were women (57%), white (81%), and under 35 years old (54%). About 15% had some prior experience working in a startup. The experimental design was identical to Study 2, with subjects still being randomly assigned to either *Male* or *Female Founders* (now omitting the

[13] Nine percent of respondents were excluded from our sample. We verified that these excluded subjects were similar to other subjects in their main characteristics, mitigating concerns that our restricted sample could be biased.

neutral condition) and then asked to count the number of persons in 12 successive photos. Upon completion, subjects were again asked whether they would be willing to code additional photos without extra pay. To test our claim that working for a female founder—even in an online setting—affects subjects' perceptions of an "ideal worker" behavior, we replaced the post-test questions by two new questions (see Appendix C). First, we asked: "*We are relatively new to mTurk, was it fair of us to ask if you could do some extra work?*". Second, to assess whether their expectations of work demands varied by founder gender, we asked: "*Was the HIT harder or easier than you expected*?." Higher values, coded on a 5-point Likert scale, indicated that subjects considered the work harder than expected. Appendix Table 11A shows that the sample is balanced in observable characteristics across the different treatments.[14]

4.2.2. Founder Gender and Workers' Expectations of Work Demands

Table 6 re-estimates the analyses from Study 2 but now includes tests for whether the effects are mediated by the respondent's expectation of work demands. If our theory is supported, we should find that respondents perceive the requests differently when confronted with a male or female founder, and that their interpretation of the request being unfair or the work demands being higher than expected determines their willingness to offer additional labor. First, we find the same results as in the first experiment: respondents are less likely to supply additional labor to female than male founders (model 1; p <0.01) and their volume of extra work is lower in that same case (model 2; p <0.001). Models 3–4 re-estimate the same specifications but now add the employee's perception of the fairness of the request. As expected, when respondents perceive the request for extra work as unfair, they are less likely to accept it (model 3; p <0.001) and they offer lower volumes of work for the same pay (model 4; p <0.001). Model 5 suggests that the odds that a request is perceived as unfair increase when subjects work for female founders (p <0.05). In models 6–8, we adopt a similar approach but focus on the respondent's perception regarding whether the work appears to be more demanding than anticipated. We find that subjects' perceptions of work difficulty are negatively correlated both with their *Likelihood of Extra Work* (model

[14] A post-hoc power analysis showed that the minimum total sample size required to detect our effect with 0.8 power was equal to 788 participants.

6; p <0.05) and with their *Extra Work Volume* (model 7; p <0.1). Model 8 further shows that the *Female Founder* treatment is positively associated with the perception that the *Work is Harder than Expected*.

*******Insert Table 6 about here*******

The next step in formally testing a mediation effect is to assess whether the indirect effects of each mediator (*Request is Unfair* and *Work is Harder than Expected*) on each outcome (*Likelihood of Extra Work* and *Extra Work Volume*) are statistically significant. We implement a common counterfactual-based approach to mediation, which estimates direct and indirect effects using a bootstrapped sample of 5,000 replications (Preacher, Rucker, and Hayes, 2007). This non-parametric test allows for greater power even in smaller samples and does not rely on normality assumptions.

Table 7 reports direct, indirect, and total effects, leading to two important conclusions. First, the indirect effect of *Female Founders* on the *Likelihood of Extra Work* via perceived unfairness is statistically significant (p=0.000 and zero falls outside the 95 percent confidence interval). The direct effect remains statistically significant, indicating partial mediation. Second, the indirect effect of *Female Founder* on *Extra Work Volume* via perceived unfairness is also statistically significant (p=0.000), while the direct effect remains significant, indicating, again, partial mediation. Together, these results suggest that the effect of perceived unfairness partially mediates the female founders' disadvantage in eliciting extra effort from workers. In terms of the perceived work difficulty, the indirect effect of *Female Founder* on the *Likelihood of Extra Work* via work perceptions is statistically significant (p=0.000), while its direct effect remains statistically significant (p=0.000), indicating partial mediation. Lastly, the indirect effect of *Female Founder* on *Extra Work Volume* via work difficulty perception is not significant (p=0.140), indicating that this perception does not mediate the effect of female founders on *Extra Work Volume*.

*******Insert Table 7 about here*******

For robustness, we performed alternative mediation tests by computing the average causal mediation effect (ACME) and assessing its sensitivity. Appendix Table 12A reports these results. Consistent with Table 7, ACME is statistically significant in all models except in panel d).[15]

Overall, we find that the perception that a request is unfair leads employees of female founders to decline this request or to provide smaller volumes of extra labor. At the same time, the perception that the work is harder than anticipated also leads participants in female-founded startups to decline requests for extra work. Overall, we find empirical support for our theory.

4.3. *Alternative Explanations for the Female Founder Disadvantage*

We have found strong support for our hypothesis in two different studies as well as evidence in line with our theoretical mechanisms in a third study. Nevertheless, to further probe our theory, we must rule out a number of competing mechanisms that may drive our findings. We report several tests based on our observational and experimental data that exclude the validity of several confounding explanations.

Lower Effort by Female Founders. One may be worried that female founders in our observational data (Study 1) work fewer hours or that they have family obligations constraining their work time, which could explain their employees' lower labor supply. Although hours worked are not available for founders in QP data (and go beyond the focus of our study), we assess if female founders were less committed than male founders as paid employees, prior to founding. Any evidence of lower commitment to wage work will likely carry over into entrepreneurship, and may trigger the same behavior among future employees. However, for founders observed in paid employment in t-1 (with t being the founding year), we find that female founders worked more, not fewer, regular hours per month than their male counterparts (108.3 versus 103.8 hours; p = 0.000) and that their weekly work schedule was also longer than men's (27 versus

[15] To address the concern that some unobserved variables may be correlated with the mediator and the outcome, we computed the parameter ρ to assess the correlation between the error terms of the models predicting the mediators and the outcome variables. For models 3 and 4 of Table 6, the ACME of the perceived unfairness could be identified when ρ varies, respectively, between $-0.25 \leq \rho \leq 1$ or $-0.26 \leq \rho \leq 1$. For models 6 and 7 of Table 6, the ACME of the perceived work difficulty could be identified when this parameter ranges between $-0.07 \leq \rho \leq 1$ or $-0.06 \leq \rho \leq 1$. In conclusion, the causal mediation effect is robust to alternative tests and sensitivity analyses.

25.6 hours; $p = 0.000$). Further, women who later became founders worked as many extra hours (0.918) as women in general in the labor market (0.920) ($p = 0.910$), while men who later became founders worked fewer extra hours (1.37) than the average male employee (2.09, $p = 0.000$).[16]

General Gender Bias. If working for a female founder creates frustration among employees, our findings of an employee labor imbalance against female founders could be explained by a broader female penalty rather than a specific resistance to work extra time for female founders. We revisit Study 2 to test whether respondents "shirk", i.e. reduce all forms of effort when working for a woman, or whether our results pertain specifically to requests for extra labor. First, we estimated a count model predicting the respondent's *Work Errors* (models 1-4, Table 13A). We computed the difference between the respondent's answer and the correct answer (i.e. the objective count of relevant items on each photo), and then summed these errors across all 12 photos. Second, we measure *Work Time* (models 5–8, Table 13A), i.e. the number of minutes dedicated to coding the 12 initial photos. Fewer work errors and longer time devoted to the job could indicate more effort by the subject. We do not find any evidence that the treatment alters the accuracy of a respondent's work (models 1–4). Although model 8 shows that respondents spent less time working for female founders ($p<0.01$) than they did for male founders, this difference did not affect the overall quality of the work, suggesting that respondents were not actively sabotaging female founders. Appendix Figures 3A and 3B depict the number of work errors, with and without controls. Figures 4A and 4B do the same for the total time spent at work. These patterns are inconsistent with gender bias being the primary explanation for respondents' behavior.

Gendered Beliefs about Startup Success. Another plausible explanation for the female founder disadvantage is that respondents presume female founders are more likely to fail and therefore conserve their time rather than spending it in a non-promising endeavor. Although these beliefs are less likely to drive the gap we identify experimentally, as respondents were temporary workers who lacked significant information about the firm's performance or prospects, we nevertheless conducted a short survey in

[16] Furthermore, we do not find any evidence that our effect is amplified for female founders in fertility age (i.e., below 45 years old), who could work less due to family constraints.

which we asked the subjects about their expectations for the firm's performance and the founder's ability to tackle challenges (cf. Appendix B). Specifically, we asked respondents whether they anticipated that the firm would face external or internal challenges and the odds that the firm would succeed. We did not find any significant association between founder's gender and their responses to any of these questions.

Gendered Expectations of Bonus. Another confounding explanation for the employee labor imbalance could be the expectation of a lower payoff from female founders. Specifically, in the mTurk setting, workers may offer extra labor hoping that it will eventually be compensated through a bonus, which employers could choose to do in this platform. If employees perceive women as less likely to provide these bonuses—because they assume that women will fail and/or are less generous—the gender of the founder could directly affect an employee's effort. Although the subjects in our experiments were *explicitly* told that there would not be a bonus for additional labor, we have still assessed whether respondents' expectations for pay varied with the founder's gender, by asking those who completed all tasks whether the pay was appropriate. Respondents were instructed to rate their pay on a 7-point scale (from way too little to way too high). We found no significant differences between treatment and control conditions (on average, respondents rated the pay as ~2.95, i.e. "a little below average").

Labor Contribution. Finally, individuals may be less likely to supply additional labor because they anticipate that the marginal value of working extra time is lower for female founders—e.g., if they believe that female founders are so unlikely to succeed that any extra work will not matter. To test this mechanism, we repeated our initial experiment (Study 2) while also manipulating the perception of an individual's contribution for each of the founders in a 2×2 design. In the low-contribution condition, workers were told that their effort will be one of the 50,000 image-tagging jobs completed for the startup. By contrast, in the high-contribution condition, workers were told that the firm needed only 500 images tagged. In the neutral condition, workers received no message about their relative contribution. If respondents adjust their behavior according to their perception of its utility, we should find a higher likelihood of working extra time when their contribution for the start-up's goal is perceived to be stronger. However, we found that, while respondents continued to exhibit a preference for male founders, there was no evidence that adjusting perceived utility affected their decision.

4.4. Summary of Results

Table 8 summarizes the three types of tests conducted: a) tests validating our hypothesis, b) tests validating the underlying mechanisms, and c) tests ruling out alternative explanations. Our results reflect the combination of two types of studies, one from extensive observational data (Study 1) and another based on two experiments (Studies 2 and 3). Both types of studies provide convergent findings that employees are more likely to decline requests for extra work (for the same pay) when the founder is a woman and that they consider a founder's gender when estimating work demands. Both research designs have strengths and weaknesses. With the observational data, we are able to document large-scale empirical patterns and thus establish external validity, but isolating the underlying mechanisms is challenging. The advantage of the experimental studies lies in their suitability for unpacking causal relationships and documenting micro-level mechanisms, which remain either difficult to tease out or unobserved in the field. Yet external validity is often compromised in the latter case. Thus, the combination of observational and experimental data has the benefits of external and internal validity.

*******Insert Table 8 about here*******

5. DISCUSSION

A well-established finding in entrepreneurship research is that female founders are subject to multiple biases that constrain their opportunities to enter entrepreneurship (Ruef et al., 2003; Aldrich, 2005; Kim et al., 2006; Yang and Aldrich, 2014). The public outcry over this has led to a wave of initiatives to remedy these biases and increase female entrepreneurship. While this is surely a noble pursuit, in this study we propose that the challenges faced by female founders do not end with the founding of their firms. Instead, we build upon work that conceptualizes employees' motivation and effort as a critical resource (Agarwal 2019; Agarwal et al., 2019), as well as research on the gendering of management and leadership roles (Cullen and Perez-Truglia, 2021; Eagly et al., 1995, Eagly and Karau, 2002) to develop a theory about a novel source of bias: one that stems from employees and their willingness to provide labor.

By pairing observational, large-scale, employer-employee linked data with two experiments, we offer evidence consistent with our theory and unpack the underlying mechanisms. First, using observational data, we found initial evidence that employees of female-founded firms contribute less aggregate time than employees of male founders, as evidenced by fewer regular and extra hours worked in female-led startups. We then leveraged two online experiments to provide additional evidence of this employee labor penalty towards female founders and to unpack the underlying mechanisms. We highlight two findings that merit particular attention. First, the experimental results were consistent with observational findings; with no differences in firms, pay, or potential benefit other than the gender of the founders, the respondents continued to provide male founders with more total labor than they did to female founders. Second, and equally critically, this labor imbalance appeared in both experiments, despite the fact that each draw was run at different times and with new respondents. Moreover, the final experiment yielded evidence that this persistent female disadvantage is, at least in part, explained by gendered assumptions about female founders and their work demands, consistent with ample research documenting pre-conceived notions that women do not prioritize work given their disproportionate family obligations (Cha, 2010; Cha and Weeden, 2014; Hochschild and Machung, 2012). In particular, we find that employees are less willing to provide extra labor to female founders because they find such requests unfair or more difficult than expected. We conduct additional analyses to rule out alternative explanations. Importantly, we found no evidence that subjects a) had different beliefs about the success of female and male founders or about their contribution to the firm’s goals, b) anticipated different bonuses from different founders, c) were more likely to shirk when working for a female founder. Our results are also unlikely to be driven by lower effort among female founders given the patterns we find in Study 1.

Overall, our multi-method empirical design establishes a solid link between a founder’s gender and an employee’s labor supply for a given pay. Our results thus support the notion that the willingness to provide labor is an important source of variation across startups and that the founder’s gender has a significant influence on employees’ motivation to commit their effort to the firm. More generally, given that resource constraints imposed upon new firms often pressure them to ask employees to contribute extra labor with no immediate pecuniary compensation, the differential receptivity to these requests constitutes a real and significant disadvantage for female entrepreneurs.

Our study offers significant contributions by extending past research on gender inequality and entrepreneurship. Whereas much scholarly attention has focused on gender disparities in founding rates (Aldrich, 2005; Castellaneta et al., 2021; Guzman and Kacperczyk, 2019; Kim et al., 2006; Ruef et al., 2003; Thébaud, 2015), much less is known about the sources of female disadvantage post-entry, when women are in charge of managing a new firm. Our study advances this scant but emerging line of work by identifying an important source of hardship: employees' provision of labor. We expand these debates by turning our attention to startup employees as an important but so far unaddressed source of disadvantage for female founders. The findings of this research highlight the critical importance of extending our theorizing beyond initial resource providers (e.g., investors) to incorporate stakeholders and audiences critical for founders once the initial hurdles are surmounted. In this regard, our study calls for greater attention to startup employees and their unequal labor supply as a key resource in growing ventures.

Our findings also extend the line of work on gendering of managerial and leadership roles (Cullen and Perez-Truglia, 2021; Eagly et al., 1995, Eagly and Karau, 2002) more generally. First, by turning our attention to startups rather than established firms—which have been the traditional focus of past research (e.g., Cullen and Perez-Truglia, 2021; Ranganathan and Shivaram, 2020; Srivastava and Sherman, 2015)—we uncover a new and so far neglected disadvantage faced by female leaders: employees' reluctance to offer them additional labor (for equal pay). Second, by documenting that female founders' requests for extra labor are perceived as unfair or harder to complete, our findings shed new light on the origins of gender bias in leadership: the unmet expectations that female leaders demand less effort in part because they tend to deprioritize work. Past studies have attributed gender bias in leadership to widely held cultural beliefs that positions of power or authority—such as management or executive roles—critically require masculine qualities (e.g., "think manager, think male"). By focusing on female founders as examples of female leaders, our findings uncover a new explanation for women's hardship as managers relative to their male counterparts and offer some evidence for its drivers. In short, we contribute to the current debates on the sources of bias against female leaders and its consequences for organizations.

Finally, we contribute to a vibrant stream on gender disparities in career outcomes (Cha & Weeden, 2014; Cotter et al., 1997; England, 1992; Huffman & Velasco, 1997; Reskin, 2000). This research has found that superiors often discriminate female subordinates; we extend these findings by documenting that subordinates can also discriminate against their female superiors. Negative stereotypes against founders may thus originate not only in upper organizational levels, but also in lower positions. Discriminatory behaviors may thus be more prevalent within firms than previously thought.

Although our study is the first, to our knowledge, to establish evidence that employee bias might put female startups at significant disadvantage post-entry, there are important limitations. Most obviously, our experiments were conducted in an online environment and used temporary workers. Much as laboratory studies are not perfect analogs to employee behavior, personal ties and organizational features surely affect the degree to which these patterns manifest in full-time employees. However, the use of a substitute population is often a necessity for experimental methods and the pertinent question is therefore whether the behavior exhibited by the temporary workers can illuminate a potential bias in their full-time counterparts. We believe that this is a suitable test for our theory for many reasons. First, even if we limit the findings to part-time employees, these differences likely affect startup performance because, given the resource limitations that startups face (Stinchcombe, 1965), sourcing labor through temporary contracts or in online settings is relatively common (Fry, 2017; Corporaal and Lehdonvirta, 2017; Kässi and Lehdonvirta, 2016). Second, an online setting offers an opportunity for random assignment, critical to establish a causal relationship between employees' willingness to accept requests for additional labor and a founder's gender. As a result, prior studies have often used the behavior of temporary employees as a window into the behavioral traits that can influence all potential employees (Burbano, 2016; Leung, 2017). Finally, while we recognize the possibility that full-time employees or onsite employees may respond differently to a founder's gender, it is noteworthy that we find similar patterns in observational registry data from Portugal. Hence, while a repeated interaction with the founders may, to some extent, de-emphasize the salience of gender, our results indicate that such bias is unlikely to be eliminated entirely. At the same time, future inquiries might want to replicate our findings within a sample of full-time workers and over a period of repeated interactions. Similarly, our research design examined employees' responsiveness to a single form of labor—repetitive, mundane tasks. However, it is plausible that employees' willingness to provide extra labor varies with the properties of the task requested. In particular, it might be that employees are even more likely to deny requests for extra labor when a task is more complex and more time-consuming. Because the costs of accepting a founder's request might be relatively smaller when tasks are mundane, our current findings may be stronger when task complexity and its associated costs increase. We present evidence that this behavior is partially driven by employees' expectations, but we cannot fully rule out that this behavior also reflects either a taste-based or a statistical form of gender discrimination. In addition, our paper presents evidence of an employee labor imbalance both in experimental settings and in observational data, but researchers may want to fruitfully identify the conditions under which such penalty can be reduced. For example, future work may benefit from assessing how additional cues about the "business gender" (e.g., whether the startup offers a product that is female or male-typed) affect employee willingness to provide extra labor. Finally, we do not offer any evidence that this labor penalty alone damages organizational performance. While it is logical to presume that this would constitute an even more serious disadvantage for female founders, our findings do not assess the degree to which this disadvantage lowers performance. Thus, we strongly encourage future research to continue to explore this area and to assess how variations in employee effort may influence organizational performance.

References

Abeler, J., A. Falk, L. Goette, D. Huffman. 2011. Reference Points and Effort Provision. *American Economic Review* 101(2) 470-92.

Abowd, J. M., Kramarz, F., Woodcock, S. (2008). Econometric analyses of linked employer-employee data. In L. Mátyás & P. Sevestre (Eds.), The econometrics of panel data 22, 727–760. Berlin: Springer.

Abraham, M., Burbano, V. 2021. Congruence between Leadership Gender and Organizational Claims Affects the Gender Composition of the Applicant Pool: Field Experimental Evidence. Organization Science, articles in advance available at https://doi.org/10.1287/orsc.2021.1442

Acker, J. 1990. Hierachies, Jobs, Bodies. *Gender & Society* 4(2) 139-158.

Agarwal, R. 2019. Human Enterprise. A. Nyberg, T. Moliterno, eds. *Handbook of Research on Strategic Human Capital*. Elgar, 1-29.

Agarwal, R., S. Braguinsky, A. Ohyama. 2019. Centers of gravity: The effect of stable shared leadership in top management teams on firm growth and industry evolution. *Strategic Management Journal*.

Aldrich, H., M. Ruef. 2006. *Organizations Evolving*, 2 ed, Thousand Oaks, CA.

Balachandra, L., Briggs, T., Eddleston, K., & Brush, C. 2019. Don't pitch like a girl!: how gender stereotypes influence investor decisions. *Entrepreneurship Theory and Practice* 43(1) 116–137

Aguinis, H., Villamor, I., & Ramani, R. S. 2020. MTurk research: Review and recommendations. Journal of Management

Baron, J.N., M. Hannan. 2002. Organizational blueprints for success in high-tech start-ups: Lessons from the Stanford project on emerging companies. *California Management Review* 44(3) 8-36.

Baron, J.N., M. Hannan, M.D. Burton. 2001. Labor pains: Change in organizational models and employee turnover in young, high-tech firms. *American Journal of Sociology* 106(4) 960-1012.

Baron, J.N., M.T. Hannan, M.D. Burton. 1999. Building the Iron Cage: Determinants of Managerial Intensity in the Early Years of Organizations. *American Sociological Review* 64(4) 527-547.

Beaman, L, Chattopadhyay, R, Duflo, E, Pande, R & Topalova, P 2009, 'Powerful women: Does exposure reduce bias?', *Quarterly Journal of Economics*, vol. 124, no. 4, pp. 1497-1540

Benard, S., S.J. Correll. 2010. Normative discrimination and the motherhood penalty. *Gender & Society* 24(5) 616-646.

Bianchi, S. M., Milkie, M. A., Sayer, L. C., & Robinson, J. P. 2000. Is anyone doing the housework? Trends in the gender division of household labor. Social Forces, 79, 191–228

Bigelow, L., L. Lundmark, J. McLean Parks, R. Wuebker. 2014. Skirting the issues: Experimental evidence of gender bias in IPO prospectus evaluations. *Journal of Management* 40(6) 1732-1759.

Bird, B.J. and Brush, C.G. 2002. A gendered perspective on organizational creation. Entrepreneurship Theory and Practice 26 (3) 41–65.

Branstetter, L., Lima, F., Taylor, L., Venâncio, A. 2014. Do Entry Regulations Deter Entrepreneurship and Job Creation? Evidence from Recent Reforms in Portugal. *The Economic Journal* 124(57) 805-832.

Brett, J.M., & Stroh, L.K. 2003. Working 61 plus hours per week: Why do managers do it? *Journal of Applied Psychology* 88(1) 67–78.

Brooks, A.W., L. Huang, S.W. Kearney, F.E. Murray. 2014. Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences* 111(12) 4427-4431.

Burbano, V.C. 2016. Social Responsibility Messages and Worker Wage Requirements: Field Experimental Evidence from Online Labor Marketplaces. *Organization Science* 27(4) 1010-1028.

Burtch, G., S. Carnahan, B.N. Greenwood. 2018. Can You Gig It? An Empirical Examination of the Gig Economy and Entrepreneurial Activity. *Management Science* 64(12) 5497-5520.

Burton, M. D., and C. M. Beckman. 2007 "Leaving a legacy: Position imprints and successor turnover in young firms." American Sociological Review, 72(2) 239–266.

Burton MD, Dahl MS, Sorenson O. 2018. Do start-ups pay less? *ILR Rev.* 71(5) 1179-1200.

Buttner, E.H., B. Rosen. 1988. Bank loan officers' perceptions of the characteristics of men, women, and successful entrepreneurs. *Journal of Business Venturing* 3(3) 249-258.

Buttner, E.H., B. Rosen. 1989. Funding new business ventures: Are decision makers biased against women entrepreneurs? *Journal of Business Venturing* 4(4) 249-261.

Byström, K., K. Järvelin. 1995. Task Complexity Affects Information Seeking and Use. *Information Processing & Management* 31(2) 191-213.

Campbell, D.J. 1988. Task Complexity: A Review and Analysis. *Academy of Management Review* 13(1) 40-52.

Campero, S., Kacperczyk, A. 2020. "Asymmetric Gender Homophily in the Startup Labor Market." Pp. 329–59 in Employee Inter- and Intra-Firm Mobility, Vol. 41, Advances in Strategic Management, edited by D. Tzabbar and B. Cirillo. Bingley, UK: Emerald Publishing Limited.

Cardon, M.S. 2003. Contingent labor as an enabler of entrepreneurial growth. *Human Resource Management* 42(4) 357-373.

Carpenter, J., Matthews, P., & Schirm, J. 2010. Tournaments and office politics: Evidence from a real effort experiment. *American Economic Review*, *100*, 504–517.

Castellaneta, F., Conti, R., Kacperczyk, A. 2020: The (Un)intended consequences of institutions lowering barriers to entrepreneurship: the impact on female workers. *Strategic Management Journal* 41(7) 1274–1304

Castilla, E. 2008. Gender, Race, and Meritocracy in Organizational Careers. *American Journal of Sociology* 113(6) 1479-1526.

Castilla, E.J., S. Benard. 2010. The paradox of meritocracy in organizations. *Administrative Science Quarterly* 55(4) 543-676.

Cha, Y. 2010. Reinforcing Separate Spheres: The Effect of Spousal Overwork on Men's and Women's Employment in Dual-Earner Households. *American Sociological Review* 75(2) 303-329.

Cha, Y. 2013. Overwork and the Persistence of Gender Segregation in Occupations. *Gender & Society* 27(2) 158-184.

Cha, Y., K.A. Weeden. 2014. Overwork and the Slow Convergence in the Gender Gap in Wages. *American Sociological Review* 79(3) 457-484.

Chapman, J.B., 1975. Comparison of male and female leadership styles. *Academy of Management Journal* 18 (3) 645–650.

Charness, G., U. Gneezy, A. Henderson. 2018. Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization* 149 74-87.

Cialdini, R.B., N.J. Goldstein. 2004. Social Influence: Compliance and Conformity. *Annual Review of Psychology* 55(1) 591-621.

Cohen, P. N., Huffman, M. L. 2007. Working for the woman? Female managers and the gender wage gap. *American Sociological Review*, 72(5): 681–704.

Coleman, S., A. Robb. 2009. A comparison of new firm financing by gender: evidence from the Kauffman Firm Survey data. *Small Business Economics* 33(4) 397.

Connelly, R. The Effect of Child Care Costs on Married Women's Labor Force Participation. *Review of Economics and Statistics*, February 1992, 74(1), pp. 83-90.

Correll, S.J. 2001. Gender and the Career Choice Process: The Role of Biased Self-Assessments. *American Journal of Sociology* 106(6) 1691-1730.

Correll, S.J., S. Benard, I. Paik. 2007. Getting a job: Is there a motherhood penalty? *American Journal of Sociology* 112(5) 1297-1338.

Cotter, David A., JoAnn DeFiore, Joan M. Hermsen, Brenda M. Kowalewski, and Reeve Vanneman. 1997. All Women Benefit: The Macro-Level Effect of Occupational Integration on Gender Earnings Inequality. *American Sociological Review* 62(5) 714–734.

Crump, J. V. McDonnell, T. M. Gureckis, 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE* 8, e57410.

Cullen, Zoë B., and Ricardo Perez-Truglia. 2021. The Old Boys' Club: Schmoozing and the Gender Gap. Working Paper.

Dahl, M.S., S. Klepper. 2015. Whom do new firms hire? *Industrial and Corporate Change* 24(4) 819-836.

DellaVigna, S., D. Pope. 2016. What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies* 85(2) 1029-1069.

DeSantola, A. Gulati, R. 2017. Scaling: Organizing and growth in entrepreneurial ventures. *Academy of Management Annals* 11(2) 640-668.

Desmond, M. 2008. *On the fireline*. University of Chicago Press.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review* 109(3) 573–598.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin* 117(1) 125–145.

Eatough, E. M., Meier, L. L., Igic, I., Elfering, A., Spector, P. E., & Semmer, N. K. (2016). You want me to do what? Two daily diary studies of illegitimate tasks and employee well-being. *Journal of Organizational Behavior* 37(1) 108–127.

England, P. 1992. *Comparable worth: Theories and evidence*. Transaction Publishers.

Gneezy, U., A. Rustichini. 2000. Pay Enough or Don't Pay at All. *The Quarterly Journal of Economics* 115(3) 791-810.

Fehr, Ernst, and Urs Fischbacher. 2004. Third-Party Punishment and Social Norms. Evolution and Human Behavior 25 (March): 63-87.

Gartside D, Silverstone Y, Farley C, Cantrell SM 2013. Tr*ends Reshaping the Future of HR: The Rise of the Extended Workforce* (Accenture, Chicago).

Gorman, E.H., and J.A. Kmec. 2009. Hierarchical rank and women's organizational mobility: Glass ceilings in corporate law firms. *American Journal of Sociology* 114(5) 1428-1474.

Greenberg, J., E. Mollick. 2017. Activist Choice Homophily and the Crowdfunding of Female Founders. *Administrative Science Quarterly* 62(2) 341-374.

Guzman, Jorge, and Aleksandra Kacperczyk. 2019. Gender Gap in Entrepreneurship. *Research Policy* 48(7) 1666-1680

Hallen, B. L. 2008. The causes and consequences of the initial network positions of new organizations: From whom do entrepreneurs receive investments? *Administrative Science Quarterly* 53(4) 685–718.

Heilman, M. E. 1980. The impact of situational factors on personnel decisions concerning women: Varying the sex composition of the applicant pool. *Organizational Behavior and Human Performance*, 26(3) 386–395.

Heilman, M.E. 1984. Information as a deterrent against sex discrimination: The effects of applicant sex and information type on preliminary employment decisions. *Organizational Behavior and Human Performance* 33(2) 174-186.

Heilman, M.E., R.F. Martell, M.C. Simon. 1988. The vagaries of sex bias: Conditions regulating the undervaluation, equivaluation, and overvaluation of female job applicants. *Organizational Behavior and Human Decision Processes* 41(1) 98-110.

Hendy, K.C., J. Liao, P. Milgram. 1997. Combining Time and Intensity Effects in Assessing Operator Information-Processing Load. *Human Factors* 39(1) 30-47.

Hietaniemi, L., Santamaria, S., Kacperczyk, A., Pltonen, J., 2021. Entrepreneurial Hiring Strategies in Standalone and Portfolio Ventures. Working Paper.

Hillman, A, Canella, AA & Harris, IC 2002. Women and racial minorities in the boardroom: How do directors differ? *Journal of Management* 28(6) 747-763.

Hochschild, A., A. Machung. 2012. *The second shift: Working families and the revolution at home*. Penguin.

Honoré, F., & Ganco, M. 2020. Entrepreneurial teams' acquisition of talent: Evidence from technology manufacturing industries using a two-sided approach. *Strategic Management Journal* in press.

Horvath, M., Herleman, H. A., & McKie, R. L. 2006. Goal orientation, task difficulty, and task interest: A multilevel analysis. *Motivation and Emotion* 30(2) 171–178.

Hsu, D. H. 2007. Experienced entrepreneurial founders, organizational capital, and venture capital funding. *Research Policy* 36(5) 722-741.

Huang, Land, Pearce, J. 2015. Managing the Unknowable: The Effectiveness of Early-Stage Investor Gut Feel in Entrepreneurial Investment Decisions. *Administrative Science Quarterly* 60(4): 634-70.

Huffman, M. L., & Velasco, S. C. 1997. When more is less: Sex composition, organizations, and earnings in US firms. *Work and Occupations* 24(2): 214–244.

Jacobs, J.A., K. Gerson. 2004. *The time divide: Work, family, and gender inequality*. Harvard University Press.

Jennings, J.E., C.G. Brush. 2013. Research on women entrepreneurs: challenges to (and from) the broader entrepreneurship literature? *Academy of Management Annals* 7(1) 663-715.

Kacperczyk, A., and Kang, S., Paik, Y., 2021. The Double-Edge Sword of a Female Political Leader: Diversity-Enhancing and Reducing Effects in the Workplace. Working Paper.

Kacperczyk, A., Marx, M. (2016). Revisiting the Small-Firm Effect on Entrepreneurship: Evidence from Firm Dissolutions. *Organization Science* 27(4) 893–910.

Kacperczyk AJ, Younkin P 2021. The founding penalty: Evidence from an audit study. *Organization Science*, forthcoming.

Kassi O and Lehdonvirta V (2016) *Online labour index: measuring the online gig economy for policy and research*. Politics & Policy 2016, Oxford, 22–23 September.

Irwin, D., and J. M. Scott. 2010. "Barriers Faced by SMEs in Raising Bank Finance." International *Journal of Entrepreneurial Behaviour & Research* 16(3): 245-259.

Kim, A., Kim, Y., Han, K., Jackson, S. E., & Ployhart, R. E. 2017. Multilevel influences on voluntary workplace green behavior: Individual differences, leader behavior, and coworker advocacy. *Journal of Management* 43(5) 1335-1358

Kim, P.H., H.E. Aldrich, L.A. Keister. 2006. Access (Not) Denied: The Impact of Financial, Human, and Cultural Capital on Entrepreneurial Entry in the United States. *Small Business Economics* 27(1) 5-22.

Leung, M.D. 2014. Dilettante or Renaissance Man? How the Order of Job Experiences Affects Hiring in an External Labor Market. *American Sociological Review* 79(1) 136-158.

Leung, M.D. 2017. Learning to Hire? Hiring as a Dynamic Experiential Learning Process in an Online Market for Contract Labor. *Management Science* 64(12) 5651-5668.

Litwin A, Phan P (2013) Quality over quantity: Reexamining the link between entrepreneurship and job creation. Indust. *Labor Relations Review* 66(4) 833–873.

Liu, P., Z. Li. 2012. Task Complexity: A Review and Conceptualization Framework. *International Journal of Industrial Ergonomics* 42(6) 553-568.

Livingston, Beth A. and Timothy A. Judge. 2008. "Emotional Responses to Work-Family Conflict: An Examination of Gender Role Orientation among Working Men and Women." *Journal of Applied Psychology* 93(1) 207-216.

Lombard, K. 2001. Female self-employment and demand for flexible, nonstandard work schedules. Economic Inquiry 39(2): 214–237.

Lussier, R. N. (1995) A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management* 33(1) 8–20.

Naumovska, I., Wernicke, G., Zajac, E. 2020. Last to come and last to go? On the complex role of gender and ethnicity in the reputational penalties for directors linked to corporate fraud. *Academy of Management Journal* 63(3) 881-902.

Nowak, M. A., Page, K. M. & Sigmund, K. 2000. Fairness versus reason in the ultimatum game. *Science* 289 1773–1775

O'Donnell, E., E.N. Johnson. 2001. The Effects of Auditor Gender and Task Complexity on Information Processing Efficiency. *International Journal of Auditing* 5(2) 91-105.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. 2007. Assessing moderated mediation hypotheses: *Theory, methods, and prescriptions. Multivariate Behavioral Research*, 42, 185–227

Quadlin, N. 2018. The mark of a woman's record: Gender and academic performance in hiring. *American Sociological Review* 83(2) 331–360.

Ranganathan, A., & Shivaram, R. (2020). Getting their hands dirty: How female managers motivate female worker productivity through subordinate scut work. *Management Science* 67(5) 3299–3320.

Reid, E. 2015. Embracing, Passing, Revealing, and the Ideal Worker Image: How People Navigate Expected and Experienced Professional Identities. *Organization Science* 26(4) 997-1017.

Reskin, B. F. 2000. The proximate causes of employment discrimination. *Contemporary Sociology* 29(2) 319-328.

Roach, M., Sauermann, H. 2015. Founder or Joiner? The Role of Preferences and Context in Shaping Different Entrepreneurkal Interests. *Management Science* 61(9) 2160-2184.

Rocha, V., van Praag, M. 2020. Mind the Gap: The Role of Gender in Entrepreneurial Career Choice and Social Influence by Founders. *Strategic Management Journal* 41(5) 841-866.

Ruef, M., H.E. Aldrich, N.M. Carter. 2003. The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs. *American Sociological Review* 68(2) 195-222.

Rudert, S. C. and Greifeneder, R., 2016. When it's okay that I don't play: Social norms and the situated construal of social exclusion. *Personality and Social Psychology Bulletin* 42(7) 955-969,

Sauermann, H. 2018. Fire in the belly? Employee motives and innovative performance in start-ups versus established firms. *Strategic Entrepreneurship Journal* 12(4) 423-454.

Semmer, N., Tschan, F., Meier, L., Facchin, S., Jacobshagen, N. 2010. Illegitimate tasks and counterproductive work behavior. *Applied Psychology: An International Review* 59(1) 70–96.

Semmer, N. K., Jacobshagen, N., Meier, L. L., Elfering, A., Beehr, T. A., Kaelin, W., & Tschan, F. 2015. Illegitimate tasks as a source of work stress. *Work & Stress* 29(1) 32-56.

Sine, W. D., Mitsuhashi, H., & Kirsch, D. A. 2006. Revisiting Burns and Stalker: Formal structure and new venture performance in emerging economic sectors. *Academy of Management Journal*, 49: 121-132.

Solal, I., & Snellman, K. 2019. Women don’t mean business? Gender penalty in board composition. *Organization Science* 30 (6) 1270–1288.

Sorenson, O., M.S. Dahl, R. Canales, M.D. Burton. 2021. Do Startup Employees Earn More in the Long Run? *Organization Science* 32(3) 587-604.

Stinchcombe, A.L. 1965. Social Structure and Organizations. J. March, ed. *Handbook of Organizations*. Rand McNally, Chicago, 153-193.

Thébaud, S. 2010. Gender and Entrepreneurship as a Career Choice. *Social Psychology Quarterly* 73(3) 288-304.

Thébaud, S. 2015a. Business as Plan B: Institutional Foundations of Gender Inequality in Entrepreneurship across 24 Industrialized Countries. *Administrative Science Quarterly* 60(4) 671-711.

Thébaud, S. 2015b. Status Beliefs and the Spirit of Capitalism: Accounting for Gender Biases in Entrepreneurship and Innovation. *Social Forces* 94(1) 61-86.

Thébaud, S., A.J. Sharkey. 2016. Unequal Hard Times: The Influence of the Great Recession on Gender Bias in Entrepreneurial Financing. *Sociological Science* 3(1) 1-31.

Tolbert, P.S., R.J. David, W.D. Sine. 2011. Studying Choice and Change: The Intersection of Institutional Theory and Entrepreneurship Research. *Organization Science* 22(5) 1332-1344.

Tonin, M., Vlassopoulos, M. 2014. An Experimental Investigation of Intrinsic Motivations for Giving. *Theory and Decision*, 76(1): 47–67.

Tuscherer, T., Sacco, D. F., Wirth, J. H., Claypool, H. M., Hugenberg, K., & Wesselmann, E. D. 2015. Responses to exclusion are moderated by its perceived fairness. *European Journal of Social Psychology* 46(3) 280-293.

Yang, T., H. Aldrich. 2014. Who’s the Boss? Explaining Gender Inequality in Entrepreneurial Teams. *American Sociological Review* 79(2) 303-327.

Yang, T., Kacperczyk, A., Naldi, L. 2021. Career Antecedents of Female Entrepreneruship. *Working Paper.*

TABLES

Table 1. Employee Labor in Female-led and Male-led Start-ups: Two-tailed *t*-tests for Full-time Employees (Study 1, Observational Data, Portugal 2002-2012)

	Female-led start-ups	Male-led start-ups	Difference
Number of regular hours per month	169.58	170.04	-0.46 ***
	(11.73)	(9.77)	
Number of extra hours per month [a]	18.67	20.12	-1.45 ***
	(17.66)	(17.27)	
Regular period of work per week	39.65	39.78	-0.13 ***
	(2.48)	(1.94)	

[a] Statistics for the subset of individuals with a positive number of extra hours. *** p < 0.001.

Table 2. Descriptive statistics: Employees in start-up firms (Study 1, Observational Data, Portugal, 2002-2012)

	All employees - all firms				Full-time employees - all firms			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Number of monthly hours worked	152.8	43.468	0	212	169.922	10.298	0	212
Number of supplementary hours worked (month)	0.701	4.82	0	173	0.742	5.045	0	173
Normal work period (week)	38.335	7.06	0	49	39.744	2.087	0	49
Female founders	0.269	0.443	0	1	0.272	0.445	0	1
Mixed gender founders	0.111	0.314	0	1	0.114	0.318	0	1
Female employee	0.409	0.492	0	1	0.414	0.493	0	1
Age	35.024	10.124	18	60	35.02	10.076	18	60
Born in Portugal	0.949	0.221	0	1	0.953	0.212	0	1
Below high school education	0.746	0.434	0	1	0.732	0.434	0	1
High school completed	0.17	0.376	0	1	0.186	0.389	0	1
University Education	0.072	0.259	0	1	0.073	0.26	0	1
Education unknown	0.009	0.099	0	1	0.009	0.095	0	1
Years in wage employment	10.742	6.282	0	23	10.814	6.283	0	23
Years as employer/entrepreneur	1.631	3.222	0	22	1.662	3.249	0	22
Monthly wage (log)	6.337	0.5	4.559	8.43	6.422	0.44	4.657	8.43
Mean age of the founder	38.079	8.996	18	60	38.301	8.992	18	60
Founder has below high school	0.598	0.479	0	1	0.58	0.479	0	1
Founder has high school completed	0.214	0.394	0	1	0.222	0.399	0	1
Founder has university education	0.169	0.363	0	1	0.179	0.372	0	1
Founder has education unknown	0.019	0.137	0	1	0.019	0.135	0	1
Founder's years in wage employment	11.809	5.822	0	23	11.889	5.814	0	23
Founder's years as employer/entrepreneur	5.081	3.697	0	22	5.121	3.736	0	22
Share of native-born founders	0.978	0.142	0	1	0.98	0.136	0	1
Firm age	3.239	1.228	1	5	3.252	1.228	1	5
Firm size (log number of employees)	2.466	1.044	0.693	6.178	2.378	0.999	0.693	6.178
Firm sales (log)	12.576	1.579	7.636	18.968	12.581	1.579	7.636	18.968
Part-time employees	0.198	0.398	0	1	0	0	0	0
Observations (employee-firm-year)	582,417				461,973			

Table 3. The Effect of Founder Gender on an Employee's Hours Worked (Study 1, Observational Data, Portugal 2002-2012)

	Multi-level/hierarchical linear models			**Count Models**		
	Number of regular hours per month (1)	Number of extra hours per month (2)	Regular period of work per week (3)	Number of regular hours per month (4)	Number of extra hours per month (5)	Regular period of work per week (6)
Founder Characteristics						
Female founders	-0.458 ***	-1.414 ***	-0.100 ***	-0.002 ***	-0.061 +	-0.002 *
	(0.051)	(0.362)	(0.011)	(0.001)	(0.036)	(0.001)
Mixed gender founders	0.049	-2.395 ***	0.013	0.000	-0.094	0.000
	(0.070)	(0.528)	(0.015)	(0.001)	(0.062)	(0.001)
Mean age of the founder	-0.020 ***	-0.019	-0.006 ***	-0.000 *	-0.000	-0.000 **
	(0.003)	(0.020)	(0.001)	(0.000)	(0.002)	(0.000)
% Founders with high school completed	0.003	-2.642 ***	-0.010	-0.000	-0.097 *	-0.001
	(0.059)	(0.447)	(0.013)	(0.001)	(0.044)	(0.001)
% Founders with university education	-0.397 ***	-0.465	-0.113 ***	-0.002 +	0.007	-0.002 *
	(0.072)	(0.474)	(0.016)	(0.001)	(0.050)	(0.001)
% Founders with unknown education	0.670 ***	1.256	0.037	0.004 **	0.084	0.001
	(0.160)	(1.167)	(0.035)	(0.001)	(0.064)	(0.001)
Founder's years in wage employment (mean)	-0.012 **	-0.008	-0.005 ***	-0.000	-0.000	-0.000 +
	(0.004)	(0.032)	(0.001)	(0.000)	(0.003)	(0.000)
Founder's years as employer/entrepreneur (mean)	0.032 ***	0.024	0.007 ***	0.000	-0.001	0.000 +
	(0.006)	(0.045)	(0.001)	(0.000)	(0.004)	(0.000)
% Native born founders	-0.059	0.095	-0.034	0.001	0.043	0.001
	(0.154)	(1.208)	(0.033)	(0.002)	(0.108)	(0.002)
Employee Characteristics						
Female employee	-0.003	-0.797 *	-0.035 ***	0.001 *	-0.048 **	0.000
	(0.046)	(0.317)	(0.010)	(0.000)	(0.017)	(0.000)
Age	-0.034 ***	0.068 **	-0.007 ***	-0.000 ***	0.003 **	-0.000 ***
	(0.003)	(0.021)	(0.001)	(0.000)	(0.001)	(0.000)
Born in Portugal	-0.297 ***	-1.023	-0.063 ***	-0.002 *	-0.057	-0.003 ***
	(0.090)	(0.772)	(0.016)	(0.001)	(0.056)	(0.001)
High school completed	-0.667 ***	0.004	-0.103 ***	-0.004 ***	-0.004	-0.005 ***
	(0.050)	(0.394)	(0.009)	(0.001)	(0.028)	(0.000)
University Education	-3.688 ***	-2.848 ***	-0.618 ***	-0.024 ***	-0.166 **	-0.022 ***
	(0.091)	(0.814)	(0.017)	(0.001)	(0.053)	(0.001)
Education unknown	0.131	3.475 *	0.020	0.001	0.125	-0.002
	(0.142)	(1.556)	(0.022)	(0.001)	(0.087)	(0.001)
Years in wage employment	0.000	0.053	0.003 **	-0.000 *	0.003 +	-0.000 *
	(0.005)	(0.036)	(0.001)	(0.000)	(0.002)	(0.000)
Years as employer/entrepreneur	-0.014 **	-0.098 *	-0.004 *	-0.000	-0.004 *	-0.000
	(0.007)	(0.048)	(0.001)	(0.000)	(0.002)	(0.000)
Monthly wage (log)	5.984 ***	0.206	0.904 ***	0.040 ***	0.032	0.036 ***
	(0.047)	(0.432)	(0.008)	(0.002)	(0.074)	(0.002)
Firm Characteristics						
Firm age	0.350 ***	0.012	0.091 ***	0.001 **	0.013	0.000
	(0.013)	(0.124)	(0.002)	(0.000)	(0.014)	(0.000)
Firm size (log no. Employees)	-0.651 ***	-1.431 ***	-0.094 ***	-0.004 ***	-0.109 ***	-0.004 ***
	(0.026)	(0.212)	(0.005)	(0.001)	(0.023)	(0.001)
Firm sales (log)	-0.031 *	0.309 **	-0.015 ***	-0.001 **	0.016	-0.001 ***
	(0.014)	(0.130)	(0.002)	(0.000)	(0.015)	(0.000)
Year FE, 2-digit Industry FE, County FE and Employee Qualification (1digit occupation) FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	461,973	17,485	461,973	461,973	17,485	461,973

+ p < 0.10; * p < 0.05; ** p < 0.01; *** p < 0.001. Models 1 to 3 are three-level models with random intercepts at both the firm and the worker-within-firm levels. These models account for the nested structure of our data, i.e., workers nested within firms. Level-3 intraclass correlation at the firm-level varies between 0.034 and 0.324 depending on the model. Level-2 intraclass correlation at the employee-within-firm varies between 0.522 and 0.792. Models 4 to 6 are count models with clustered standard errors at the firm-level. Models 4 and 6 are Poisson models, while Model 5 is a Negative Binomial model, which fits the data better due to the overdispersion in the dependent variable. The goodness of fit of each model is assessed with a Likelihood Ratio test comparing Poisson and Negative Binomial models for each dependent variable. Models (2) and (5) are restricted to employees reporting a positive number of extra hours. Table 5A reports a Zero-Inflated Negative Binomial model, which allows to include the zeros in the estimation. All the models above are restricted to full-time employees. Table 7A reports the key coefficients when including part-time employees and controlling for type of contract.

Table 4. Respondent Characteristics by Treatment: Balance Table (Study 2, Experimental Data)

Panel A: Male Founder vs. Control Condition						
	Control (Unnamed Founders: N=104)		**Treatment (Male Founders: N= 171)**			
	Mean	**SD**	**Mean**	**SD**	**Diff**	**p-value**
Male Respondent	0.470	0.500	0.470	0.501	0.001	0.989
White	0.727	0.446	0.655	0.477	0.072	0.103
Education	3.507	0.828	3.524	0.965	0.032	0.854
Worked for Startup	0.157	0.365	0.115	0.321	0.042	0.315
Age	2.450	1.104	2.878	3.835	-0.427	0.088
Panel B: Female Founder vs. Control Condition						
	Control (Unnamed Founders: N=104)		**Treatment (Female Founders: N=182)**			
	Mean	**SD**	**Mean**	**SD**	**Diff**	**p-value**
Male Respondent	0.485	0.501	0.448	0.499	0.038	0.435
White	0.705	0.457	0.692	0.463	0.013	0.765
Education	3.526	0.058	3.494	0.825	0.032	0.716
Worked for Startup	0.111	0.314	0.200	0.402	-0.089	0.025
Age	2.693	3.027	2.461	1.091	0.232	0.342

Table 5. The Impact of Founder's Gender on Employee Effort (Study 2, Experimental Data)

	Likelihood of Extra Work				**Extra Work Volume**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
<u>Treatments</u>								
Male Founders	1.139*** (0.276)		2.354*** (0.338)		1.064*** (0.116)		1.159*** (0.120)	
Female Founders	0.750** (0.275)	-0.389+ (0.215)	1.453*** (0.304)	-0.899** (0.304)	0.604*** (0.121)	-0.460*** (0.076)	0.937*** (0.122)	-0.221** (0.084)
<u>Control Variables: Respondent's Characteristics</u>								
Male			-0.700** (0.253)	-0.489 (0.298)			-0.263** (0.081)	-0.117 (0.087)
Education			0.344* (0.141)	0.431** (0.166)			0.086* (0.042)	0.096* (0.044)
White			-0.574 (0.305)	-0.698 (0.384)			-0.280** (0.089)	-0.270** (0.095)
Worked for Startup			0.111 (0.355)	0.031 (0.398)			0.001 (0.107)	0.019 (0.112)
Age			-0.047 (0.051)	-0.055 (0.063)			0.082 (0.005)	0.085*** (0.005)
Constant	-1.150*** (0.229)	-0.011 (0.152)	-1.530* (0.654)	0.559 (0.757)	-0.145 (0.111)	0.919*** (0.048)	-0.322 (0.213)	0.723*** (0.211)
Specification	Logit	Logit	Logit	Logit	Poisson	Poisson	Poisson	Poisson
Observations	457	353	333	229	457	353	333	229
Pseudo R^2	0.030	0.007	0.166	0.078	0.041	0.016	0.378	0.386

Standard errors in parentheses; + p < 0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; two-tailed tests

Table 6. The Impact of Founder's Gender on Employee Effort: Mediating Effect of Unfairness and Task Difficulty (Experimental Data, Study 3)

	Likelihood of Extra Work	**Extra Work Volume**	**Likelihood of Extra Work**	**Extra Work Volume**	**Request is Unfair**	**Likelihood of Extra Work**	**Extra Work Volume**	**Work is Harder than Expected**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Female Founders	-0.282**	-0.189***	-0.226	-0.137**	0.275*	-0.249+	-0.219+	0.190***
	(0.139)	(0.056)	(0.144)	(0.056)	(0.140)	(0.140)	(0.127)	(0.052)
Male Respondent	-0.129	-0.136**	-0.089	-0.102+	0.173	-0.131	-0.170	-0.018
	(0.141)	(0.057)	(0.146)	(0.057)	(0.140)	(0.141)	(0.128)	(0.052)
Education	0.317	0.190**	0.245	-0.036	-0.036	0.312	0.289	-0.034
	(0.210)	(0.080)	(0.217)	(0.234)	(0.234)	(0.210)	(0.198)	(0.081)
White	-0.188	-0.117*	-0.178	-0.106	0.111	-0.182	-0.155	0.016
	(0.176)	(0.070)	(0.182)	(0.070)	(0.179)	(0.177)	(0.161)	(0.066)
Ever Worked at Startup	-0.041	-0.032	0.033	0.012	0.276	-0.056	-0.063	-0.073
	(0.195)	(0.079)	(0.203)	(0.078)	(0.190)	(0.196)	(0.178)	(0.073)
Age	0.492***	0.307***	0.503***	0.293***	-0.093	0.486***	0.433***	-0.040
	(0.103)	(0.039)	(0.106)	(0.039)	(0.107)	(0.103)	(0.095)	(0.039)
Request is Unfair			-1.393***	-1.145***				
			(0.183)	(0.085)				
Work is Harder than Expected						-0.181*	-0.146+	
						(0.086)	(0.078)	
Constant	-1.523***	-0.201	-1176***	0.071	-0.607+	-1.144**	0.944**	
	(0.332)	(0.129)	(0.345)	(0.129)	(0.352)	(0.377)	(0.350)	
Specification	Logit	Poisson	Logit	Poisson	Logit	Logit	Poisson	Linear
Observations	982	982	982	982	982	982	982	982
Log Likelihood	-601.5	-1968	-567.7	-1850	-603.05	-599.3	-1965	
Adj. R-Squared								0.011

Note: Robust standard errors in parenthesis, +*p*<.10, **p*<.05, ***p*<.01, ****p*<.001 two-tailed tests; no neutral condition included.

Table 7. Mediation Analysis: Direct, Indirect and Total Effects (based on results in Table 6)

	Estimate	Bootstrapped Standard Error	*p*-value	95% Conf. Interval	
a) Likelihood of Extra Work mediated by "Request is Unfair"					
Natural direct effect	0.79	0.12	0.001	0.59	1.05
Natural indirect effect	0.94	0.03	0.000	0.89	0.99
Total effect	0.75	0.12	0.000	0.56	1.00
b) Extra Work Volume mediated by "Request is Unfair"					
Natural direct effect	0.87	0.08	0.000	0.73	1.05
Natural indirect effect	0.95	0.02	0.000	0.90	0.99
Total effect	0.83	0.08	0.000	0.90	1.00
c) Likelihood of Extra Work mediated by "Work is Harder than Expected"					
Natural direct effect	0.78	0.11	0.000	0.59	1.03
Natural indirect effect	0.97	0.02	0.000	0.92	0.10
Total effect	0.75	0.11	0.000	0.57	0.99
d) Likelihood of Extra Work Volume mediated by "Work is Harder than Expected"					
Natural direct effect	-0.22	0.13	0.090	-0.46	0.03
Natural indirect effect	-0.03	0.02	0.140	-0.08	0.00
Total effect	-0.25	0.13	0.050	-0.49	0.00

Table 8. Summary of Empirical Tests

Testing Hypothesis 1	
Observational Data (Study 1)	
Baseline Findings	Table 1 (raw data); Table 3 (estimations)
Robustness checks based on model specification	Appendix Tables 4A, 5A, and 7A
Robustness checks based on different sampling choices	Appendix Tables 6A and 9A
Experimental Data (Study 2)	
Baseline Findings	Table 5; Table 6 (models 1-2); Appendix Figures 1A-2B
Testing the Underlying Mechanisms (Study 3)	
Perceptions of unfairness	Table 6 (models 3-5); Table 8 (panels a and b); Appendix Table 12A (panels a and b)
Perceptions of work difficulty	Table 6 (models 6-8); Table 8 (panels c and d); Appendix Table 12A (panels c and d)
Tests Ruling Out Alternative Explanations	
Differences in male and female founders' effort	section 4.3
General gender bias	Appendix Table 13A; Appendix Figures 3A-4B
Gendered beliefs about startup success	section 4.3
Different expectations of bonus by founder's gender	section 4.3
Perceptions of labor contribution	section 4.3

Figure 1: Control and Treatment Conditions

Initial Prompt:

We recently founded a photo-based startup and need help coding these photographs for our database.

We should need help on a few projects, the first of which involves the identification of people on our images. Your help is meaningful as these projects are core to our company's services. MTurkers' efforts will contribute to about 15% of our output for today. The task is described below.

Please count the number of people in each image. Include items in the background, and items who are partially or mostly hidden from view.

Control	Male Founders Treatment	Female Founders Treatment
	Matthew & Joe, Co-Founders Photolytics	Amanda & Chloe, Co-Founders Photolytics

Second Prompt:

This is all the work we can presently pay you for, but if you have some extra time, we have a few more pictures we could use help coding.

Would you be willing to code a few more photos? You can stop at any time.

Control	Male Founders Treatment	Female Founders Treatment
	Matthew & Joe, Co-Founders Photolytics	Amanda & Chloe, Co-Founders Photolytics