

Double standards?

Co-authorship and gender bias in early stage academic hiring

Klarita Gërxhani, European University Institute¹²

Nevena Kulic, University of Florence³

Fabienne Liechti, University of Lausanne

Abstract

This article studies the hiring intentions of tenured professors with regard to early career researchers. In particular, it examines gender bias in evaluations of academic hiring in Italy and investigates whether this bias depends on collaborative work and its related conventions across academic fields. We rely on status characteristics theory to test our hypotheses via a factorial survey (vignette) experiment of 2,098 associate and full professors employed in Italian public universities in 2019. This is one of the few experiments of the hiring process in academia conducted on a nationally representative population of university professors. Our article focuses specifically on three academic fields: humanities, economics, and social sciences. The results indicate that female academics in Italy are penalized for co-authoring. They receive less favorable evaluations of their qualifications, but only when the evaluator is a man. As hypothesized, this gender bias is found in economics, a field where conventions of co-authorship allow for more uncertainty about individual contributions to a joint publication.

Keywords: Gender bias, academia, factorial survey experiment, co-authorships, disciplines

¹ The order of the authors is alphabetical. All authors contributed equally to this article.

² The authors acknowledge funding from the Research Council of the European University Institute.

³ Corresponding author: Nevena Kulic, University of Florence, email: nevena.kulic@unifi.it

Introduction

One of the major questions when studying gender inequality in academia is whether there is a systematic bias in the evaluation of women's academic record and suitability for an academic career. Evaluation may relate to various aspects of academic activity, such as research, teaching, networking, etc. In particular, there are concerns about whether the same standards are applied when evaluating the scientific production of women and men. Publications are considered a strong indication of academic performance in academia, and it has been theoretically argued and empirically shown that returns to academic collaborations are different for women and men. For instance, male political scientists who collaborate at the same rate as women, profit more from these collaborations, as they are able to translate them faster into measurable scientific product: they produce, on average, two publications more than women, so co-authorships tend to ultimately amplify gender differences in publications (Djupe et al., 2019). Moreover, women receive less professional recognition than men for their collaborative work, even when they contribute equally or more to these collaborations (Abramo et al., 2013). The phenomenon of women's overlooked contributions to collaborative undertakings is often termed the "Matilda effect" after the name of the scientist who first formulated it (Lincoln et al., 2012). In a recent empirical study, Sarsons et al. (2021) found that after controlling for the quality of scientific work, male economists in the US are tenured at a similar rate regardless of whether they have co-authored or single-authored publications, while women are less likely to obtain tenure in economics the more they co-author.

The underlying argument of gender differentiated returns to collaborative work is that when evaluating a person's publication record, the uncertainty of who contributed what is much higher with a co-authored publication than with a single-authored publication. This is more unfavorable for women than men if there are expectations that women are less able or less committed than men. To better understand what

happens in such high-uncertainty situations, status characteristics theory and its arguments about women's competence and their contribution to group work may prove to be valuable (Correll et al., 2007, Ridgeway, 2009). This theory predicts that women will receive less credit than men for collaborative work as long as there are generally held beliefs that women in general, and female academics in particular, have either lower levels of competence or are less committed than men. The limited available empirical evidence (Sarsons et al., 2021) supports these expectations, but only for the field of economics and in the US context. Given that different academic disciplines operate based on different institutional contexts (Pontille, 2003) and on different conventions (Waltman, 2012), the question arises as to whether such biased behavior can be found across academic fields and outside the US. After all, if co-authorship does indeed send different signals for women than for men, this could help address gendered hiring and promotion patterns in the academic world. This argument is particularly relevant today as collaborative work in many disciplines, including social sciences, is increasing (Abramo et al., 2013). Hence, the following questions underlie the research presented in this article. Does bias occur in the evaluation of equally productive men and women in academia, and what is the role therein of the composition of their scientific output in terms of single and co-authored publications? Does the extent of biased evaluations of co-authored work vary across disciplines?

The article addresses these questions using a factorial survey experiment (also known as vignette experiment). In such experiments, respondents are asked to evaluate fictional candidates based on their descriptions that vary on several characteristics (Auspurg and Hinz, 2015). Factorial survey experiments have been widely employed for studying hiring behaviour (Damelang and Abraham, 2016; Di Stasio, 2014; Di Stasio and Gërkhani, 2015; Liechti, 2019; McDonald, 2017). In contrast to observational data, studying hiring behaviour by experimental data has the advantage that the researchers can control for the

information provided to participants and therefore limit potential bias due to unobservable characteristics. To our knowledge, this is one of the few experiments for hiring in academia based on a nationally representative population of university professors. Limited previous research has studied gender bias using observational and administrative data (Sarsons, 2017; Quadlin, 2018). A very recent study uses a survey experiment to examine possible gender bias in the process of academic hiring in the Nordic context (Carlsson et al., 2020). Their results do not show any such bias against equally qualified candidates to associate professor positions. However, considering the “persistent gender gap in academia in the Nordic region” (p. 1), they conclude that their study “does not rule out the possibility that men experience advantages in other phases of academic life, such as in monitoring, review boards, or peer-review assessments.” (p. 9). By focusing on gender differentiated returns to *collaborative work*, our study contributes to the unravelling of these ‘other phases’ which could be hidden channels of gender bias.

In the experiment, we focus on the following disciplines in the Italian academiaⁱ: history, geography, psychology, pedagogy and philosophy (broadly denoted as “the humanities” henceforth), economics, and social sciences. The main reason for doing so is the relatively homogeneous nature of their evaluating criteria, which enables comparisons across these disciplines based on the same experiment.ⁱⁱ Our factorial survey experiment aims to mimic the decision-making process that is assumed to take place when hiring candidates for assistant professorships -with and without tenure-track features- in the Italian academic setting. We chose to focus on this crucial moment in an academic career because that path is characterized by a notable lack of women in tenured or tenure track positions succeeding doctoral studies. This phenomenon is known as the ‘leaky pipeline’ (van Anders, 2004). The vignette experiment was conducted in June 2019 among associate and full professors at public Italian universities. These participants were asked to assess two inter-related dimensions of the potential academic success of

fictitious candidates: (1) the probability that they would be shortlisted (invited for an interview), and (2) how qualified they were for the position. Capturing both dimensions enables a better understanding of the findings: the former dimension reflects potential behavioral action in the form of shortlisting, while the latter reflects the evaluator's beliefs about the candidate's qualification for the job.

Gender as a status characteristic and co-authorship signals in academia

Status characteristics theory posits that there are different evaluations and perceptions of individuals as a result of their status characteristics, such as gender, race and social class (Ridgeway, 2019). When gender is used as a status characteristic, one category of a nominal distinction (i.e. men) is associated with greater competence, i.e., higher status, than the other category (i.e. women). Gender discrimination is, therefore, driven by a cognitive bias that incorporates widely held cultural beliefs about the worthiness and competence of individuals associated with each gender (Correll, 2006; Correll and Benard, 2006; Ridgeway and Krichelin-Katz, 2013). This cognitive bias becomes particularly salient in mixed-gender environments in which men's and women's levels of competence or commitment are ambiguous. An extension of status characteristics theory, known as the 'double standard approach', argues that even when competence is known and is equivalent for men and women, those with a perceived lower status (i.e. women) will be evaluated based on stricter standards than those with a higher status (i.e. men) (Foschi et al., 2019). "As a result, equal task performances are more likely to be judged as indicative of ability when performed by a higher status member of the group." (Correll and Ridgeway, 2003:41).

Academic jobs are culturally considered a male occupation (Coate and Howson, 2016; Witz, 1990), substantiated by scant female representation at higher professorship levels. This is similar to what is seen in high level management positions (van den Brink et al. ,2016; Reskin and Bielby, 2005). Thereby, discriminatory behavior towards female academics may result from status beliefs concerning their lesser

ability, their lower level of commitment, or due to stricter evaluation standards (Correll and Benard, 2006; Ridgeway, 2009). Having an academic career implies an expectation about the high level of ability and commitment needed to succeed in the profession (Van den Brink and Benschop, 2012). One of the strong signals of academics' ability and commitment are their scientific publications. The theoretical and empirical research discussed above argues that beyond their quantity and quality, academics' scientific publications are valued differently depending on the author composition of their scientific output. In other words, signals of one's academic ability and commitment seem to vary when their scientific publications are single authored or co-authored. While single-authorship of a scientific output conveys a clear signal of its author's ability and commitment, co-authorship 'hides' the exact contribution of each single author (Sarsons et al., 2021). Therefore, compared to single-authorships, co-authored scientific publications allow for more uncertainty in evaluating how able or committed each contributing author is. This uncertainty will, in turn, allow for more evaluation bias if there are cultural beliefs about women being expected to be less able or less committed than men in a male-dominated profession like academia.

As a result, female academics may be evaluated less favorably or by stricter standards for their co-authored work, even when female and male academics are equal in terms of the quantity and quality of their scientific output. Put differently, the hiring evaluation of a male and female academic might be a function of the author composition of their scientific output, which further interacts with gender as a status characteristic. Hence, we expect *female academics to receive less favorable evaluations for their co-authored publications than their male colleagues (everything else being equal)(H1)*.

A scenario where female and male academics produce commensurate academic output (e.g., quantity and quality of publications), sending a clear signal of women and men being equally able and committed, would render gender irrelevant as a status characteristic for the evaluation. As mentioned above, this

outcome may be different when evaluating female and male academics based on their collaborative work. Status characteristics theory argues that, particularly when applied to collective tasks (i.e., members of a group working on a collective goal like a co-authored publication), women's contributions are less valued than are men's (Correll and Ridgeway, 2003; Correll, 2006). As a result, gender differentiated standards in the evaluation of men and women may be more marked, the higher the degree of uncertainty about individual contributions to a collective task (Correll and Ridgeway, 2003; Castilla, 2008; Gorman and Kmec, 2007; Foschi, 1996). This uncertainty is particularly high when name ordering of the authors is done alphabetically, without specifying the contribution of each individual author (Zuckerman, 1968; Maciejovsky et al., 2008). Non-alphabetical name ordering, that is, according to one's contribution, is considered more equal or 'merit-based' (Van Praag and Van Praag, 2008). A number of studies show that the convention on the use of one or the other type of name ordering varies across disciplines. Economics and the fields closest to it, such as finance and mathematics, seem to score the highest in the use of alphabetical name ordering of co-authored publications.ⁱⁱⁱ Disciplines like sociology, psychology, and humanities, on the other hand, score much lower in the use of the alphabetical convention (for empirical evidence, see Engers et al., 1999, Laband and Tollison, 2000, Joseph et al., 2005, Van Praag and van Praag, 2008, Waltman, 2012, Kuld and O'Hagan, 2018, Weber, 2018).

Considering the variation in conventions of name ordering of co-authored publications, that is, considering the disciplinary differences in degrees of uncertainty of who contributed what to a joint academic output (i.e., with economics scoring the highest), we expect that *the unfavorable evaluations for women's co-authored publications will be found in economics but not in humanities and social sciences (H2)*.

Case study: Entry-level hiring in Italy

We test the above hypotheses by examining whether hiring decisions at the entry-level academic positions are gender biased. We focus on the Italian case. With 36 percent women in academic employment, Italy's figures are slightly above the EU-28 average ratio of 33 percent of female academics. According to the Italian Ministry of Education, University and Research (MIUR) (own elaboration of the 2018 administrative data), women account for 22.34 percent of full professorships (slightly below the 24 percent found in the EU; European Commission, 2018), while women's share of full professorships in the social sciences and humanities is at 28 percent. These figures stand in contrast to a balanced share of women in 2019 among Italian graduates (57.6 percent) and Phd holders (51.8 percent) (Morana and Sagramora, 2019).

Academic markets used to differ widely across countries in the standards for appointments and academic duties. Following a broader transformation in the globalized European academic labour market, there have been some signs of standardization in the most recent years in Italy in the form of more regulated internal labour market (Mussolin, 2004, 2005). Italy has recently adopted a series of reforms relating to academic recruitment and career progression (Murgia and Poggio, 2019). The reforms, that took place in 2012 ("Gelmini Law" (210/2010)), resulted in the establishment of tenure track and non-tenure track assistant professorships. The position of tenure track professorships is based on the rule that after three years the employment continues into a promotion to associate professorship, dependent on the candidate having a positive evaluation and upon the receipt of a national habilitation. The national habilitation (tenure) criteria was introduced to select pre-candidates for associate and full professorships, and disadvantage those candidates that scored poorly on criterias of scientific output. The concept of habilitation as part of tenure is known across academic systems (see Mussolin, 2005 for France and Germany). However, in many countries such as the UK, the Netherlands and Germany, the tenure decision

is autonomously made by universities themselves^{iv} (Marini, 2017). The habilitation recognition in Italy is based on quantitative and qualitative assessment of one's output, with a particular focus on the number of publications that are needed for being eligible for a tenure-track position; note that this threshold differs for each scientific discipline. In contrast, the non-tenure track positions automatically end after five years of employment (three years plus an additional two, conditional on a positive evaluation) and typically do not depend on having obtained a national habilitation.

Yet, despite the formal reforms, the national habilitation has not brought to academic hiring the fairness and quality of the expected levels (Marzolla, 2015). According to Abramo et al. (2015), the Italian academic recruitment is known for a strong preference for internal candidates, which in turn sets the entry barriers into the system much higher for outsiders and young scholars in comparison to those within the departments. They argue that informal practices of 'clientelism' still dominate today and, among other things, they foster gender discrimination. In other words, it is harder for a female academic to be successful. For instance, in a study of recruitment dynamics in an Italian research institute, Checchi et al. (2019) find that the lower network density among female candidates is also a selection instrument discriminating against women. Obtaining a professorship within the Italian academic labour market, therefore, often depends on adopting the right networking strategies and on the structure and the character of individual connections (Marini and Maschitti, 2018).

As described in detail below, these main characteristics of the Italian academic system informed our vignette experiment. Particularly, the differentiation between tenure track and non-tenure track positions, the accepted standards of quality as required for the national habilitation, and the networking patterns were explicitly considered in designing the experiment.

Method

Investigating hiring decisions in academia is not an easy task. Even when the secondary data is available, the danger of not being able to control for all of the confounding factors can compromise the findings. That is why we rely on a factorial survey (vignette) experiment, which, as mentioned above, is an established method of studying employers' hiring strategies.

Our factorial survey experiment in the Italian academia

In our experiment, the participants (i.e. associate and full professors) were asked to evaluate descriptions of fictional candidates for non-tenure track and tenure track assistant professorship positions. Only associate and full professors are eligible to hold a seat in the appointment committees. Therefore, they are considered an optimal choice for the experiment due to their expected expertise in the regular decision making process in academic hiring. Each respondent evaluates profiles in their own field areas. In addition to introducing two types of early career positions, we varied the candidates' descriptions according to a number of other dimensions that help us test our theoretical hypotheses. First, those that best capture our main explanatory variables, namely candidate's gender and the author composition of their scientific publications (i.e. all co-authored vs. mostly co-authored vs. mostly single-authored), as well as whether or not the candidate had taken parental (maternity/paternity) leave (i.e. parent or not) and the status level of the candidate's ties (i.e. high vs. low).

Table 1 shows a detailed description of all the dimensions (see the appendix for an example of a vignette). Each dimension was carefully selected and described to adequately reflect the above mentioned characteristics of the Italian academic system. More specifically, while the descriptions of candidates vary in the type of publication record in terms of single-authored and co-authored work, the total number of publications was fixed throughout all the vignettes. For tenure track positions it was ten and for non-tenure track positions it was five. The total number of publications was decided on the basis of current

regulations for obtaining a position; in line with the requirement for the national habilitation, the application for tenure track positions requires typically a higher number of publications than an application for non-tenure track positions (e.g. currently, it is expected that candidates submit from ten to thirteen publications to qualify for national habilitation). The dimension of parental (maternity/paternity) leave stipulates that candidates had or had not taken mandatory maternity or paternity leave of five months (following the current state level regulation). Although parental leave is accounted for in the national habilitation, the assumption was that it could additionally impact the development of a candidate's academic career due to motherhood penalty (Correll et al., 2007) and was therefore included in the description of the vignettes. Finally, the status of ties is a proxy for the pattern of candidate's connections, that is, whether the candidate usually collaborates with academics of similar academic level or with scholars of high reputation. This dimension was included in line with the literature reporting double standards in the evaluation of network connections of men and women (Ibarra 1993, Burt, 1998; Lutter, 2015). The combination of all dimensions yields a vignette universe of 24 different combinations. Since the number of combinations is too large to be evaluated by every single participant, we grouped our vignette universe into four blocks of six vignettes each, relying on a d-optimal blocking that allows for the estimation of unconfounded main- and two-way-interaction effects (Kuhfeld, 1994, Su and Steiner, 2018). D-optimal blocking allows controlling for block effects in the empirical analysis and ensures that none of the dimensions are confounded with the blocks. Furthermore, to avoid socially desirable answers and to hide our interest in studying gender bias, we introduced gender as a between-participant dimension, meaning that half of our participants saw only female candidates and the other half, only male candidates. Such a between-participant design, in which the sensitive dimensions vary only between groups, is a common way to deal with social desirability biases in vignette studies (Walzenbach, 2019).

Table 1: Dimensions of candidates’ scientific profile included in the vignettes

<i>Dimension</i>		<i>Level</i>
Gender (between-participants) ¹		Female Male
Type of position		Tenure track position Non-tenure track position
Type of publication record	Tenure track position	Mostly single authored (7 single vs. 3 co-authored) Mostly co-authored (7 co-authored vs. 3 single) All co-authored
	Non-tenure track position	Mostly single authored (4 single vs. 1 co-authored) Mostly co-authored (4 co-authored vs. 1 single) All co-authored
Parental leave		Yes/No parental leave interruption of five months
Status of ties		High: research collaborations with influential scholars Low: research collaborations with scholars at the same academic level
<p><i>Notes:</i></p> <p>1. Gender was signalled by choosing typical Italian male and female names.</p> <p>2. The total number of publications was adopted in relation to the position for which the candidate was applying, with a higher number of publications for tenure track positions. The total number of publications was ten for tenure track positions and five for non-tenure track positions.</p>		

Vignette blocks were randomly assigned to participants. Once respondents started participating in the survey, each of them was assigned to one block. In addition, the order of the vignettes was randomized within each block to avoid order effects (Su and Steiner, 2018). The vignettes were presented sequentially but the participants were able to go back and forth between the vignettes to avoid censoring^v. Each participant was asked to evaluate a set of six candidates on the basis of two dependent variables on a scale from 1 (worst evaluation) to 7 (best evaluation). The dependent variables were: (1) the likelihood that the participant would invite the candidate for a job interview – ‘being shortlisted’, and (2) how well qualified the candidate was for the position – ‘being qualified’.

After the vignettes, additional questions were asked, concerning, for instance, participants' sociodemographic characteristics, their departments, and specifically, their departments' hiring procedures.

Data

Our sample consists of a random subsample of Italian professors (associate and full professors) in humanities, economics, and social sciences. The Ministry of Education, Universities and Research (MIUR) has compiled an official list of disciplines, which are divided into 14 major disciplinary areas, covering more than 150 disciplines. To give an example, philosophy and history are grouped together in area 11, all forms of law are defined as area 12, whereas mathematics, physics and chemistry are all in area 1. More specifically, area 11, which we broadly denote as 'humanities', includes history, geography, psychology, pedagogy and philosophy. Area 13 'economics' includes demography, economics and statistics. Area 14 -'social sciences'- includes social and political science as well as international relations. We restricted our sample to these three areas (i.e., areas 11, 13 and 14), because of the relatively homogeneous nature of their evaluating criteria, including the requirements for the national habilitation. For instance, the three areas rely mostly on non-bibliometric evaluation based on quantity of publications rather than on author's total citations and they require similar number of high quality publications needed to qualify for the habilitation ("fascia A"). This allowed us to conduct the same vignette experiment across more than one discipline. Nevertheless, these areas differ regarding the representation of women at higher academic levels. According to our own calculations from the 2018 MIUR data, around 47 percent of all associate professors in the "Humanities" are women, in comparison to 39 percent of all associate professors both in the social sciences and economics. Female professors account for 35 percent of total

full professors in humanities, and 23 and 27 percent in economics and social sciences, respectively. Overall, humanities appear to be the most gender balanced whereas economics lags most behind.

We obtained a list with all associate and full professors in Italy in April 2018 from the official database of MIUR. Out of 5,985 professors in areas 11, 13 and 14, we drew a random subsample of 2,262 individuals, stratified by rank and discipline. The email addresses of all professors in this subsample were collected manually. Of these, 164 email addresses were invalid and therefore our sample consists of 2,098 individuals. The vignettes were submitted to the participants via an online survey administered in Qualtrics® and participants were incentivized through the possibility of participating in a prize draw for an iPad. After one week, a reminder was sent to those who had not responded. The survey was online for three weeks and answering took between 15 and 20 minutes. We obtained a response rate of 18.82 percent, which resulted in 395 respondents and 2,382 rated vignettes. This response rate is comparable to those of other vignette studies with a similar population (Damelang and Abraham, 2016; Liechti, 2019).

A common challenge when conducting surveys is that there might be a response bias: respondents may differ significantly from non-respondents, thus leading to biased results. Our data allows us to partially address this bias by comparing some characteristics of respondents and non-respondents. Since we know the gender, rank, and discipline area of respondents and non-respondents, we compared the two groups with respect to these variables. The results are shown in Table A1, in the appendix. We note that women are slightly over-represented among respondents, accounting for 41.24 percent of the total, versus 36.18 percent among non-respondents. Also, associate professors and professors from social sciences were more likely to participate in the survey.

Model

Because our unit of analysis is a vignette that is evaluated by a single respondent and because each respondent has rated multiple vignettes, we estimate linear multi-level models to investigate the effects of the vignette dimensions on the two dependent variables. A multi-level framework is needed because the experiment includes two types of independent variables, one at the level of the respondent and another, at the vignette level. The dependent variable is on a scale from 1-7, therefore linear models are employed. The major independent variables of interest are the gender of the fictional candidates and the type of publication record, namely, whether the candidate's publications are only co-authored publications, mostly co-authored publications, or mostly single-authored. The models additionally control for whether the candidate has taken parental leave, the status of candidates' ties within their professional networks, and the type of position (tenure track versus non-tenure track assistant professorships), as these are characteristics that may play an important role in the hiring process.

Finally, the models control for respondents' characteristics such as gender, the disciplinary areas (humanities, economics or social sciences), rank (associate or full professor), and the affiliation to a department (departmental fixed effects). Controlling for the respective departments of the respondents allows us to compare the evaluation scores of individuals from the same departments. Because the distribution of the dependent variables is somewhat skewed to the left, ordered multi-level random intercept logistic regression was used as a robustness check. This confirms the results of the linear multi-level random intercept regression. The models were run on the whole sample and on two subsamples. One subsample is based on respondents' gender, as there might be differences in the evaluation of men and women depending on the gender of the evaluator (Bagues et al., 2017; Sarsons et al., 2021). The other subsample is based on the disciplinary areas in which respondents work, which allows us to test our

hypothesis that there is gender bias to co-authored work in economics but not in humanities and social sciences.

Results

Descriptives

A first examination of the dependent variables, displayed in Table 2, shows that the mean of being shortlisted is relatively high, namely 5.75 percent for the whole sample. There is some difference in evaluation across respondents' gender, but it is not statistically significant. Overall, the evaluation of the candidates' qualification for the position is lower than evaluations of their chances of being shortlisted, namely 5.25 vs. 5.75, again with no differences across respondents' gender. There is, however, substantial difference across disciplinary areas, with the lowest evaluations of qualification for the position and the lowest probability in shortlisting occurring in the humanities, and the highest, in economics. The difference is statistically significant, which indicates some variation in the evaluation of the candidates across different academic fields (p-value =0.00 for both outcomes).

Table 2: Mean and standard deviation of dependent variables, by participants' disciplinary area and gender

	Shortlisted		Qualified	
	Mean	Sd	Mean	Sd
All sample	5.75	1.35	5.25	1.37
Female respondents	5.80	1.29	5.25	1.33
Male respondents	5.72	1.39	5.25	1.39
Area 11 'Humanities' (history, geography, psychology, pedagogy and philosophy)	5.57	1.45	5.07	1.42
Area 13 'Economics' (demography, economics and statistics)	5.89	1.38	5.42	1.38
Area 14 'Social Sciences' (social and political science, and international relations)	5.76	1.2	5.23	1.29

Male and female candidates receive similar marks from those who evaluate them for both shortlisting and qualification for the position (analysis available upon request). In other words, when other characteristics of the candidates' profiles are not controlled for, there doesn't appear to be any gender bias in the evaluation of candidates. Also, the evaluation is not dependent on the gender of the evaluator.

Stricter standards for co-authoring female academics?

Table 3 shows the results of multi-level linear regressions - for the overall sample and for the subsample according to respondents' gender - on both dependent variables: probability of being shortlisted and judged to be qualified for the position. We first discuss the results for the whole sample (1st and 2nd columns of Table 3).^{vi} They indicate that the type of publication record does not seem to affect either of

the two dependent variables differently for male and female candidates. In other words, the interaction terms, between the candidate's gender and whether the publications are all co-authored, mostly co-authored or mostly single authored, are not significant. Therefore, based on the whole sample of respondents and controlling for all other characteristics, we do not find support for our first hypothesis, that with similar objective characteristics, female academics receive less favorable evaluations for their co-authored publications than their male colleagues. Put differently, the standards for assessing candidates for an early-career academic position in Italy are no stricter for co-authoring women than for co-authoring men.

Some of the control variables have a significant influence on the outcomes. Namely, those applying for a tenure-track position are more likely to be shortlisted and considered qualified than those applying for a non-tenure track position; candidates who collaborate with scholars of similar academic level, i.e., those who have low status ties, are evaluated more negatively in both shortlisting and being considered qualified for the position than candidates who collaborate with influential scholars, i.e., those who have high status ties. These findings are in line with the literature that reports that both the type of the position to which one applies and the reputation status of candidate's personal academic connections provide alternative signals of that candidate's academic quality (e.g., Lin et al. 1981). Put differently, the higher the status of the candidates' desired position and of their academic connections, the higher is their perceived academic quality. Having taken a period of parental leave appeared not to influence the probability to be shortlisted, but it signalled a higher qualification for the position, although of a moderate effect (0.085, significant at 1 percent level). In other words, equally productive candidates are considered more qualified if they have reached their current academic level in addition to caring for children. The inclusion of the second level variables such as discipline areas, respondent's gender and departmental

fixed effects only marginally changes the results. Moreover, additional regressions were run, controlling for respondents' total number of publications, number of single-authored publications, and the number of their co-authors. The results remain largely unchanged, but the sample size decreases by 200 observations. The only significant and positive effect found is for the number of respondents' co-authors, which positively affects the evaluation of vignettes; the more respondents co-author, the more generous they are when evaluating the vignette profiles. These additional results are available upon request.

Table 3: Multi-level linear regressions with random intercept: Likelihood of being shortlisted and of being judged qualified for a position (whole sample and subsample according to respondents' gender)

VARIABLES	<i>Whole sample</i>		<i>Male respondents</i>		<i>Female respondents</i>	
	Being shortlisted	Being qualified	Being shortlisted	Being qualified	Being shortlisted	Being qualified
<i>Vignette dimensions*</i>						
Tenure track position	0.268*** (0.034)	0.257*** (0.039)	0.216*** (0.045)	0.170*** (0.049)	0.337*** (0.050)	0.377*** (0.062)
Female candidate	-0.180 (0.147)	-0.090 (0.144)	-0.139 (0.197)	-0.041 (0.186)	-0.308 (0.224)	-0.464** (0.214)
Co-authored publications only	-0.200*** (0.057)	-0.342*** (0.066)	-0.191** (0.075)	-0.252*** (0.082)	-0.214** (0.087)	-0.477*** (0.108)
Mostly single authored publications	0.324*** (0.057)	0.381*** (0.066)	0.373*** (0.075)	0.467*** (0.082)	0.251*** (0.087)	0.251** (0.108)
Female candidate # co-authored publications only	-0.040 (0.082)	-0.098 (0.095)	-0.117 (0.111)	-0.319*** (0.121)	0.061 (0.123)	0.200 (0.152)

Female candidate # mostly single authored publications	-0.012 (0.083)	-0.004 (0.095)	-0.065 (0.111)	-0.120 (0.121)	0.067 (0.123)	0.164 (0.152)
Parental leave	0.022 (0.034)	0.085** (0.039)	-0.010 (0.045)	0.098** (0.049)	0.064 (0.050)	0.068 (0.062)
Low status ties	-0.074** (0.034)	-0.104*** (0.039)	-0.090** (0.045)	-0.131*** (0.049)	-0.052 (0.050)	-0.065 (0.062)
<i>Respondents' characteristics</i>						
Area 13 'Economics'	-0.130 (0.317)	-0.193 (0.303)	-0.059 (0.497)	-0.186 (0.459)	-0.478 (0.386)	-0.485 (0.355)
Area 14 'Social science'	-0.255 (0.232)	-0.246 (0.222)	0.076 (0.277)	0.152 (0.256)	-0.638 (0.392)	-0.893** (0.360)
Full professor	-0.076 (0.152)	-0.149 (0.145)	0.167 (0.221)	0.198 (0.204)	0.144 (0.217)	-0.026 (0.200)
Male respondent	-0.072 (0.141)	-0.170 (0.135)				
Departments	X	X	X	X	X	X
Observations	1.678	1.678	967	967	711	711
Number of groups (respondents)	289	289	167	167	122	122

Standard errors in
parentheses

Notes: (*) Reference categories are: Non-tenure track positions, male candidates, mostly co-authored publications, no parental leave, high status ties, Area 11 'Humanities', associate professor, female respondent. *** p<0.01, ** p<0.05, * p<0.1

However, the full sample may hide heterogeneous patterns between two subgroups: male and female respondents. As shown in columns three to seven of Table 3, male and female respondents do not differ in their evaluation of a candidate's probability of being shortlisted, irrespective of the candidate's type of publication record. When it comes to evaluating how qualified a candidate is considered for the position, male respondents judge female candidates whose publications are all co-authored less qualified than the equivalent male candidates. More specifically, the interaction between being a female candidate and having all publications co-authored is associated with a decrease in the evaluation of that candidate's qualification for the position by -0.316 (statistically significant at 1 percent level) on a scale of 1-7. In other words, male respondents do seem to hold stricter standards for collaborative female candidates than for collaborative male academics, net of all other characteristics. This is not the case for the subsample of female respondents. The difference between male and female respondents is statistically significant.

These effects remain unchanged after including the control variables. While the direct significant effect of some of the control variables, like tenure track position, is similar to that found for the whole sample, the effect of other controls differs across the overall sample and the two subsamples. For instance, only male respondents evaluate candidates with low status ties more negatively than candidates with high status ties, in both shortlisting and qualification for the position. Also, only male respondents evaluate candidates who are parents as more qualified than those who are not. Female respondents, on the other hand, judge a candidate to be less qualified when she is a woman whose publications are mostly co-authored and when the disciplinary area concerned is social sciences (significant at 5 percent level).

Double evaluation standards in economics?

To test our second hypothesis, which predicts unfavorable evaluations for females' co-authored publications in economics only, we ran further analyses, as reported in Table 4. We perform the analyses for the three disciplinary areas separately: area 11 'humanities', area 13 'economics', and area 14 'social sciences'. As discussed above, although applying comparable evaluating criteria, these areas differ in their conventions of how co-authors are listed in a joint publication. This varies from a listing in alphabetical order such as that practiced in economics to a based-on-individual-contribution listing, which is more common in humanities and social sciences. As a result, the degree of uncertainty in who did what in a joint publication is highest in economics.

Table 4 reports the results for different disciplinary areas for the sample of male respondents only. The results for female respondents (available upon request) do not differ qualitatively from those reported above with respect to the whole sample of female respondents. Male respondents do not differentiate between the scientific publications of male and female candidates when evaluating their likelihood to be shortlisted in all three disciplinary areas. It is when male respondents evaluate whether the candidate is qualified for the position that we observe a stricter standard towards women, but only in economics. Male economists seem to penalize female candidates who have only co-authored publications by -0.512 on a 7 point scale (significant at 1 percent) compared to men who also have only co-authored publications. Since it has been argued that economics is the discipline with the highest degrees of uncertainty regarding contributions to joint publications, this finding supports our second hypothesis. However, it does so partially, because it only holds when male economists are the evaluators.^{vii}As for the control variables, candidates applying for a tenure-track position are likely to be shortlisted and considered to be qualified for the position in all three disciplinary areas. Having low status ties negatively affects the likelihood of

being shortlisted and of being considered qualified in social sciences only. Other controls reveal a similar direction as in the previous models but show insignificant effects.

Table 4: Multi-level linear regressions with random intercept. Likelihood of being shortlisted and of being judged qualified for a position (sample of male respondents across disciplines)

VARIABLES	Being shortlisted			Being qualified		
	Human	Econ	Social	Human	Econ	Social
Tenure-track position	0.403***	-0.055	0.325***	0.367***	-0.009	0.197**
	-0.087	-0.076	-0.07	-0.088	-0.08	-0.082
Female candidate	0.809**	-0.283	-0.379	0.227	0.016	-0.377
	-0.327	-0.363	-0.248	-0.37	-0.294	-0.267
Co-authored publications only	-0.179	-0.121	-0.263**	-0.108	-0.185	-0.403***
	-0.146	-0.126	-0.115	-0.148	-0.133	-0.135
Mostly single authored publications	0.449***	0.270**	0.428***	0.518***	0.422***	0.487***
	-0.146	-0.126	-0.116	-0.148	-0.133	-0.136
Female candidate # co-authored publications only	0.021	-0.2	-0.132	-0.287	-0.512***	-0.178
	-0.212	-0.186	-0.172	-0.215	-0.197	-0.202
Female candidate # mostly single authored publications	0.051	0.018	-0.233	-0.018	-0.143	-0.189
	-0.212	-0.186	-0.172	-0.215	-0.197	-0.202
Parental leave	-0.022	-0.002	-0.015	0.047	0.115	0.115
	-0.087	-0.076	-0.07	-0.088	-0.08	-0.082
Low status ties	-0.049	-0.027	-0.167**	-0.026	-0.098	-0.225***
	-0.087	-0.076	-0.07	-0.088	-0.08	-0.082

Second level controls	X	X	X	X	X	X
Observations	245	328	394	245	328	394
Number of groups (respondents)	43	56	68	43	56	68

Notes: Reference categories as above. Second level controls include ranks and department fixed effects.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Conclusion

Despite the fact that the share of women and men at the doctoral level has been balanced for a while, the paucity of women represented in higher stages of academic career (European Commission 2015) can be considered dramatic. The mechanisms underlying this imbalance can be supply or demand driven and the two are often difficult to disentangle. This article aims to address this difficulty by isolating and uncovering some of the demand-side mechanisms and thus focuses on the gatekeeper effect underlying unequal gender representation in academia. It does so by examining whether evaluations of hiring propensities and of being considered qualified with regard to scientific output are gender biased for the entry level academic positions. If so, this bias would influence underrepresentation of women in the later stages of the academic career.

In line with status characteristics theory, we predicted that female academics would be evaluated less favorably for their collaborative work in comparison to male academics, and particularly in disciplines where it is hardest to establish the extent of individual contributions to joint publications. This was tested

by means of a factorial survey experiment among Italian professors in the humanities, economics and social sciences.

Our results show that the evaluation of academic output of Italian female academics is subject to some 'hidden' gender bias. When individual academic output becomes ambiguous, as is the case in collaborative work, gender as a status characteristic becomes activated. We find that, during the early career stage, women in Italy are indeed penalized for having all co-authored publications. Compared to their male colleagues, they receive less favorable evaluations of being considered qualified for a tenure track or non-tenure track assistant professorship, but only when the evaluator (an associate or full professor) is a man. This gender bias seems to be present in economics, a discipline which is argued to rely predominantly on conventions of co-authorships that allow for more uncertainty on individual contributions to a joint publication. Our findings complement recent observational evidence which focuses primarily on the field of economics in the US (Sarsons et al. 2021). By focusing on various academic disciplines that operate according to different co-authorship conventions, and in a different context, our experimental study contributes to a broader question of the 'boundaries' of gender bias. The larger implication of these findings is that, when it comes to collaborative work, which is ever increasing in academia, women's contribution is underestimated, confirming some of the postulates of the status characteristic theory. It is, however, important to emphasize that, contrary to the empirical evidence that finds that both sexes are likely to show bias against women when hiring in academia (Moss-Racusin et al., 2012; Bagues et al., 2017), our study underlines the presence of male bias in the Italian academic context. This result calls for raising awareness to counter such biases and to reinforce equal representation of men and women in academic organizations. A more explicit action in this direction would be to impose a quota on the gender composition of the appointing committees and thus fix the gender composition within ranks.

Interestingly, we do not observe any gender bias in the evaluators' likelihood of inviting a candidate for a job interview (i.e. shortlisting). Women are shortlisted as often as men, irrespective of the gender of the evaluator and of the discipline. This is in line with Bagues et al. (2017) who find no gender bias in getting the national habilitation, which to some extent resembles to being invited for an interview (Marini and Meschitti, 2018). On the one hand it can be expected that if a female candidate is judged to be qualified, she will also more likely be shortlisted. On the other hand, a female candidate may be judged to be less qualified and still be shortlisted, if there are internal and external pressures to increase gender equality. Such pressures are more and more present worldwide, but also within the Italian academia (MIUR, 2018). Moreover, being invited for an interview is only the first step to hiring, and inviting a 'less qualified' female candidate could signal compliance with the pressures while allowing for her exclusion in the later stages of hiring process.

Although our research partially confirms our hypotheses, we acknowledge other possibilities for the interpretation of our results. One is related to the inter-disciplinary difference in the gender composition of the respondents' academic fields. As mentioned previously, in Italy, social science and humanities fields contain larger numbers of women faculty members than economics does. For this reason, respondents from the former two fields are likely to know better the actual ability of women in their departments; that is, to know that the women perform at least as well as the men. This knowledge may have contributed to a decrease in the uncertainty of who did what in a joint publication, thus attributing equal credit to male and female academics in social science and humanities. Furthermore, the lack of bias when the evaluator is a woman could also be explained by familiarity; that is, female academics are more likely to know the actual abilities and contributions of their female colleagues.

A potential limitation of our study is that even though current evidence shows that vignette studies are able to capture a real life behavior (Hainmueller et al. 2015), participants in factorial surveys know that their answers have no real consequences and that they might still act differently in real life situations. Furthermore, by addressing the social desirability bias with a split-sample design (respondents only saw one gender, either male or female candidates), we were not able to rely on within respondent information to test how a single respondent evaluated different genders. This way we traded within respondent variability for lower social desirability bias. Another limitation is that our vignette setup does not allow for the distinction between co-authored publications with ‘friends or academic tribes’ and ‘any other scholar’. Further research is needed to unravel whether our main finding that co-authorships matter depends on such a distinction. Also, our vignettes focus on selected variables of theoretical interest and could not fully cover all the potentially relevant dimensions of an evaluation and hiring process. While the selection was necessary to be able to maintain the accuracy and the feasibility of the factorial survey experiment, it did come at a cost of evaluating a candidate under limited information. Moreover, our experiment relies on decisions of single participants and therefore underplays the importance of group interactions between committee members, which can also drive the dynamics in evaluation and hiring. Let us emphasize, however, that we do find hidden biases emerging even under the above mentioned limitations. Follow-up research can build on the contributions of our study and establish their salience in a richer context.

Literature

- Abramo, G., D'Angelo, C.A., Rosati, F. (2015) 'The Determinants of Academic Career Advancement: Evidence from Italy', *Science and Public Policy* 42(6): 761-774.
- Abramo, G., D'Angelo, C. A., Murgia, G. (2013). Gender differences in research collaboration. *Journal of Informetrics*, 7(4), 811-822.
- Auspurg, K., Hinz, T. (2015). *Factorial Survey Experiments*. Los Angeles: SAGE.
- Bagues, M., Sylos-Labini, M., Zinovyeva, N. (2017). Does the gender composition of scientific committees matter?. *American Economic Review*, 107(4), 1207-38.
- Burt, R.S. (1998). 'The Gender of Social Capital' *Rationality and Society* 10(1):5–46.
- Castilla, E. J. (2008) 'Gender, Race, and Meritocracy in Organizational Careers', *American Journal of Sociology*, 113 (6): 1479–1526.
- Carlsson, M., Finseraas, H., Midtbøen, A. H., & Rafnsdóttir, G. L. (2020). Gender Bias in Academic Recruitment? Evidence from a Survey Experiment in the Nordic Region. *European Sociological Review*: 1–12. doi: 10.1093/esr/jcaa050
- Checchi, D., Cicognani, S., Kulic, N. (2019) 'Gender quotas or girls' networks? Towards an understanding of recruitment in the research profession in Italy', *Work, employment and society*.
- Coate, K., Howson, C. K. (2016). Indicators of esteem: Gender and prestige in academic work. *British Journal of Sociology of Education*, 37(4), 567-585.

- Correll S.J., Benard, S., Paik, I. (2007) 'Getting a job: Is there a motherhood penalty?', *American Journal of Sociology*, 112(5), 1297-1338.
- Correll, S. J., Ridgeway, C. L. (2003) 'Expectation States Theory'. In Delamater, J. (ed) *Handbook of Social Psychology*, New York: Kluwer Academic/Plenum Publishers, pp. 29–51.
- Damelang, A., Abraham, M. (2016). You Can Take Some of It with You!: A Vignette Study on the Acceptance of Foreign Vocational Certificates and Ethnic Inequality in the German Labor Market. *Zeitschrift für Soziologie*, 45(2), 91-106
- Di Stasio, V. (2014). Education as a Signal of trainability: results from a vignette study with Italian employers. *European Sociological Review*, 30(6), 796–809.
- Di Stasio, V., Gërkhani, K. (2015). Employers' social contacts and their hiring behavior in a factorial survey. *Social Science Research*, 51, 93–107.
- Djupe, P. A., Smith, A. E., Sokhey, A. E. (2019). Explaining Gender in the Journals: How Submission Practices Affect Publication Patterns in Political Science. *PS: Political Science & Politics*, 52(1), 71-77.
- European Commission (2018) *She Figures 2018. Gender in Research and Innovation*. Luxembourg: Publications Office of the European Union.
- Foschi, M. (1996) 'Double Standards in the Evaluation of Men and Women', *Social Psychology Quarterly*, 59(3): 237-254.
- Foschi, M., Ndobo, A., Faure, A. (2019) 'Assessing and Blocking Double Standards for Competence', *Advances in Group Processes* 36: 19-45.

- Gorman, E. H., Kmec, J. A. (2007) 'We (Have to) Try Harder: Gender and Required Work Effort in Britain and the United States', *Gender and Society*, 21 (6): 828–856.
- Hainmueller, J., Hangartner, D., Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behaviour. *Proceedings of the National Academy of Science of the United States of America*, 112(8), 2395–2400.
- Ibarra, H. (1993). Personal networks of women and minorities in management: A conceptual framework. *Academy of Management Review*, 18(1), 56-87.
- Kuhfeld, W. F., Tobias, R. D., Garratt, M., Tobias, D. (1994). Efficient Experimental Design Applications with Marketing Research Applications. *Journal of Marketing Research*, 31(4), 545–557.
- Kuld, L., O'Hagan, J. (2018). Rise of multi-authored papers in economics: Demise of the 'lone star' and why?. *Scientometrics*, 114(3), 1207-1225.
- Liechti, F. (2019). Connecting Employers and Workers: Can Recommendations from the Public Employment Service Act as a Substitute for Social Contacts? *Work, Employment and Society*, online first. <https://doi.org/10.1177/0950017019836888>
- Lin, N., Ensel, W. M, Vaughn, J. C. (1981) 'Social Resources and Strength of Ties : Structural Factors in Occupational Status Attainment', *American Sociological Review* 46 (4): 393–405.
- Lincoln, A. E., Pincus, S., Koster, J. B., Leboy, P. S. (2012). The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s. *Social studies of science*, 42(2), 307-320.
- Lutter, M. (2015) 'Do Women Suffer from Network Closure? The Moderating Effect of Social Capital on Gender Inequality in a Project-Based Labor Market, 1929 to 2010' *American Sociological Review* 80(2): 329-358.

- Marini, G. (2017). New promotion patterns in Italian universities: Less seniority and more productivity? Data from ASN. *Higher Education*, 73(2), 189-205.
- Marini, G., Meschitti, V. (2018). The trench warfare of gender discrimination: evidence from academic promotions to full professor in Italy. *Scientometrics*, 115(2), 989-1006.
- McDonald, P. (2017). Improving our Understanding of Employer Decision-making Thanks to Factorial Survey Analysis. *LIVES Working Paper*, 75, 1-38. <http://dx.doi.org/10.12682/lives.2296-1658.2017.61>
- Miur, 2018. Indicazioni per azioni positive del miur sui temi di genere nell'università e nella ricerca, accessed 03/04/2020.
- Morana MT., Sagramora S. (2019). Focus "Le carriere femminili in ambito accademico", MIUR
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the national academy of sciences*, 109(41), 16474-16479.
- Murgia, A. & Poggio, B. (eds) 2019 *Gender & Precarious Research Careers*, London: Routledge
- Musselin, C. (2004). Towards a European academic labour market? Some lessons drawn from empirical studies on academic mobility. *Higher Education*, 48(1), 55–78.
- Musselin, C. (2005). European academic labor markets in transition. *Higher Education*, 49(1-2), 135-154.
- Pontille, D. (2003). Authorship Practices and Institutional Contexts in Sociology: Elements for a Comparison of the United States and France. *Science, Technology, & Human Values*, 28(2): 217-243.

- Quadlin, N. (2018). The mark of a woman's record: Gender and academic performance in hiring. *American Sociological Review*, 83(2), 331-360.
- Reskin, B. F., Bielby, D. D. (2005). A sociological perspective on gender and career outcomes. *Journal of Economic Perspectives*, 19(1), 71-86.
- Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. Oxford University Press.
- Ridgeway C. L. (2009) 'Framed before we know it: How gender shapes social relations', *Gender & Society*, 23(2): 145-160.
- Ridgeway, C. L. (2019). *Status: Why Is It Everywhere? Why Does It Matter?*. Russell Sage Foundation.
- Ridgeway, C. L., Kricheli-Katz, T. (2013) 'Intersecting cultural beliefs in social relations: Gender, race, and class binds and freedoms', *Gender & Society*, 27(3): 294-318.
- Ridgeway, C., Correll, S. J. (2004) 'Motherhood as a Status Characteristics', *Journal of Social Issues*, 60(4): 683-700.
- Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review*, 107(5), 141-45.
- Sarsons, H., Gërkhani, K., E. Reuben, A. Schram (2021). Gender Differences in Recognition for Group Work. Forthcoming in the *Journal of Political Economy*.
- Su, D., Steiner P. (2018). An Evaluation of Experimental Designs for Constructing Vignette Sets in Factorial Surveys. *Sociological Methods & Research*. Online First. <https://doi.org/10.1177/0049124117746427>

- Van den Brink, M., Holgersson, C., Linghag, S., & Deé, S. (2016). Inflating and down playing strengths and weaknesses—Practicing gender in the evaluation of potential managers and partners. *Scandinavian Journal of Management*, 32(1), 20-32.
- Van den Brink, M., C. Holgersson, S. Linghag (2016). Inflating and downplaying strengths and weaknesses. Practicing gender in the evaluation of potential managers and partners. *Scandinavian Journal of Management*, 32, 1, 20-32
- Van den Brink, M., Benschop, Y. (2012). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*, 19(4), 507-524.
- Van Praag, C. M., & Van Praag, B. M. (2008). The benefits of being economics professor A (rather than Z). *Economica*, 75(300), 782-796. Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, 6(4), 700-711.

ⁱ A detailed motivation of why we focus on the Italian academia is provided below.

ⁱⁱ Note that the differences on such criteria between humanities, economic and social sciences and the rest of the disciplines (such as Stem, Technical, Life sciences) is quite large, for instance regarding the propensity of publication, publication standards, or co-authorships. Different evaluation criteria are also established by the national habilitation procedure in Italy (Morzolla, 2015). For instance, the evaluation in the three disciplines under study is based on non-bibliometric indicators (e.g. quantity of publications) in contrast to hard sciences where bibliometric indicators are in place (e.g. total citations). Such differences would not allow for a proper comparison based on the same experiment.

ⁱⁱⁱ Why the use of this convention is dominant in economics, even after receiving a lot of criticism for being discriminatory (Weber 2018), remains still unclear.

^{iv} Even though the habilitation sets the common standards for faculty appointments and signals eligibility for career progression, each opening for professorships is individually advertised and guided by an appointed committee, which typically consists of several members (associate and full professors), some of whom are external to the department.

^v Censored responses refer to a situation when respondents have already chosen the end point of a limited response scale and are therefore no longer able to select an even more extreme answer for subsequent vignettes (Auspurg and Hinz 2015:124).

^{vi} In the main text, we focus primarily on the models in which there is an interaction between gender and the type of publication record. In the models in which the candidate's gender and his or her type of publication records are included separately -as main effects- gender does not appear to have any role in the evaluation process.

^{vii} Note that this finding is not generalizable to a new sample from the same population, because the difference between disciplines with regards to evaluation of women's co-authored work is not statistically significant.

Appendix

Table A1: Respondents' and Non-Respondents' characteristics

Variable	Respondents	Non-Respondents
Gender		
Female	41.24	36.18
Male	54.30	63.82
Type of Rank		
Full time Professor:	30.75	36.29
Associate Professor:	69.25	63.71
Discipline		
Economics:	32.30	41.84
Social Sciences:	41.34	25.19
Humanities:	26.36	32.98
Age	Mean:54.99 Sde: 8.57 Min: 30 Max: 72	
Length of academic career (number of years in academia)	Mean: 26.91 Sde: 8.60 Min: 1 Max:49	
Length of employment within the department (number of years)	Mean:15.02 Sde: 8.98 Min: 0.5 Max: 46	
Number of publications	Mean: 66.53 Sde: 28.10 Min: 0 Max: 250	

Table A2: Vignette set-up

An example of a vignette (dimensions varied are indicated in italics)

Introductory text:

Vignette example:

There is an opening for a *tenure track position* of an assistant professorship.

Giorgia Bianchi has a PhD from an Italian university and 6 years of Postdoc experience. She has obtained national grants for conducting her research, which has resulted in 10 peer-reviewed publications: *7 single authored and 3 co-authored*. Moreover, she has established research collaborations with *scholars of similar academic level* and has accumulated sufficient teaching experience at the Bachelor and Master levels.

Please rate the candidate on a scale from 0-7

- a) How likely it is that you would invite the candidate for a job interview
- b) How well is the candidate qualified to the position?