



## UvA-DARE (Digital Academic Repository)

### Gender inequality in 'cum laude' distinctions for PhD students

Bol, T.

**DOI**

[10.31235/osf.io/s5b6j](https://doi.org/10.31235/osf.io/s5b6j)

**Publication date**

2023

**Document Version**

Final published version

**License**

CC BY-ND

[Link to publication](#)

**Citation for published version (APA):**

Bol, T. (2023). *Gender inequality in 'cum laude' distinctions for PhD students*. (1 ed.) SocArXiv. <https://doi.org/10.31235/osf.io/s5b6j>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Gender inequality in ‘cum laude’ distinctions for PhD students

Author: Thijs Bol<sup>1,2\*</sup>

<sup>1</sup>Department of Sociology, University of Amsterdam, Amsterdam, the Netherlands

<sup>2</sup>Amsterdam Centre for Inequality Studies, University of Amsterdam, Amsterdam, the Netherlands

5 \*Correspondence to: t.bol@uva.nl

## Abstract

10 Resource allocation in academia is highly skewed, and peer evaluation is the main method used to distribute scarce resources. A large literature documents gender inequality, and the explanation for this inequality is homophily: male evaluators give more favorable ratings to male candidates. We investigate this by focusing on ‘cum laude’ distinctions for PhD students in the Netherlands, a distinction that is only awarded to 5 percent of all dissertations and has as its sole goal to distinguish the top from the rest. Using data from over 5,000 PhD recipients of a large Dutch university for the period 2011-2021, we find that female PhD students were almost two times less likely to get a ‘cum laude’ distinction than their male counterparts, even when they had the same doctoral advisor. This gender gap is largest when dissertations are evaluated by all-15 male committees and decreases as evaluation committees include more female members.

## Keywords

20 Gender inequality, science, excellence, evaluation, cum laude

## INTRODUCTION

Resource allocation in science is highly skewed (1–3). Acceptance rates at high-prestige journals are often below ten percent, and the same holds true for prestigious research grants (4).

Consequently, only a small fraction of all researchers can publish in high-prestige journals, receive research grants, or obtain tenured positions. Peer evaluation is key to determining who receives these resources: scholars review each other's work and determine who belongs to the top and who does not (5).

While peer evaluation is the dominant method to divide scarce resources, it is not uncontested. A large literature has investigated gender inequality in peer evaluation across a multitude of settings. Many studies find evidence for inequality: journal peer review scores are lower for women (6–8), men receive higher scores for their grant proposals (9–11), teaching evaluations are less favorable for women (12–15), and men are more likely to be evaluated positively in academic hiring (16, 17). The literature is not uniform in this finding, with some recent studies, for example, finding no gender inequality in journal peer review (18, 19) or academic hiring (20). However, gender inequality in evaluation is argued to be part of the explanation for gender differences in academic careers (21).

Why do female scholars receive lower evaluations than their male counterparts? Quality could be a first argument, but several studies show that differences remain even when quality is kept constant (11, 15, 22). Another important argument lies in the subjective nature of evaluations: evaluations are never objective and cannot be dissociated from evaluators (5, 23). Evaluators' conception of academic excellence is not neutral but gendered (27, 28), and they are more likely to perceive men as top achievers. Recent studies also find evidence for gender bias in group work, where women receive less credit for their contributions (22).

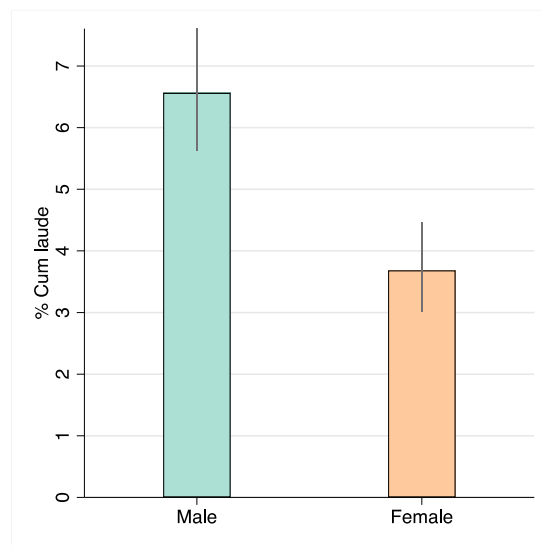
Homophily between evaluator and evaluated is also argued to be one of the key reasons for gender inequality in evaluations (8, 18, 24, 25). Particularly in contexts where there are no neutral or universal criteria that distinguish the top from the rest, evaluators are more favorable about those that are similar to themselves (23). Since the likelihood to evaluate colleagues (e.g., in hiring committees or grant panels) increases with seniority, the strong overrepresentation of men amongst senior scholars (26, 27) could reproduce existing gender inequalities when male evaluators are more likely to evaluate male academics more positively.

In this article, we study gender inequality in evaluation at the very start of the academic career, by looking at cum laude distinctions for PhD students in the Netherlands. After defending their dissertation, PhD students either obtain a regular doctoral degree or a doctoral degree with a 'cum laude' distinction. Obtaining this distinction is rare: only about 5% of all PhD students are awarded a 'cum laude'. The main doctoral advisor (always a full professor) initiates the procedure to award a 'cum laude' (see *SM Appendix* for a detailed description of the procedure). A separate dissertation committee consisting of five to seven scholars with relevant expertise then decides. The committee must unanimously agree that the dissertation should be awarded 'cum laude'. If the committee votes for 'cum laude', two external referees (that are provided by the doctoral advisor) must attest this to be the case. If they all believe the dissertation to rank

among the top 5%, a cum laude is awarded. PhD students have no influence on this process and are not informed. Only after defending their dissertation, they will learn whether their dissertation was awarded a ‘cum laude’ distinction. Candidates who do not receive the award will generally never learn whether a ‘cum laude’ procedure was initiated.

This empirical case provides an interesting context to study gender inequality in evaluations for three reasons. First, there are no clear criteria that define a ‘cum laude’ distinction—this all depends on what evaluators perceive to be important. This absence of any formal criteria (see *SM Appendix* for more information) thereby provides a most-likely case to observe gender inequality in evaluations. It also allows us to study the role of homophily: most doctoral advisors and members of doctoral committees are male, and we investigate whether male evaluators more likely to evaluate male PhD students as excellent. Second, the only purpose of a ‘cum laude’ is to separate the exceptional from the rest. By creating an artificial scarcity (5%), the awarding of cum laude distinctions is one of the clearest manifestations of the search for excellence that permeates academia (28, 29). Finally, since most doctoral advisors have supervised multiple PhD students over the period of observation (2010-2021), we can exploit fixed effects models (30) where we study gender differences in the likelihood of obtaining ‘cum laude’ or homophily effects of the dissertation committee within the same doctoral advisor. This allows us to cancel out explanations that point to between-advisor variation, for example in supervision quality, or their likelihood to apply for ‘cum laude’.

For our analyses, we rely on the information of 5,239 PhD students that received their doctorate between 2011 and 2021 at a Dutch public university that is one of the largest research universities in Europe. All PhD students that have defended their dissertations in the mentioned period are part of the data. The university under study covers all scientific domains and disciplines.

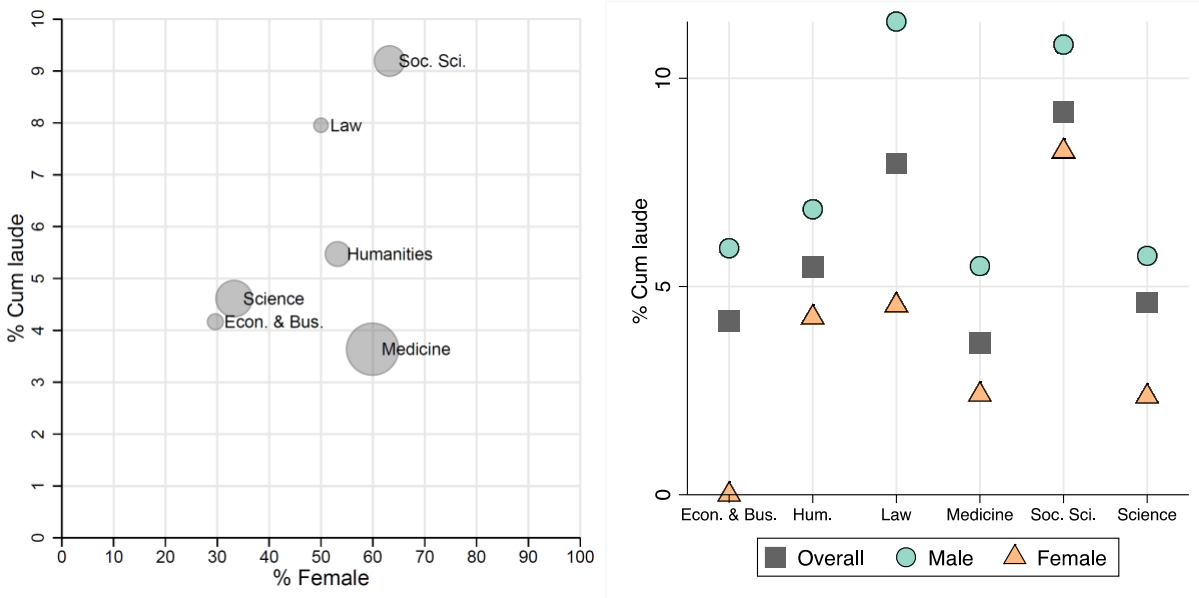


**Fig. 1. Gender gap in ‘cum laude’ distinctions.** The average gender difference in awarded cum laude distinctions for PhD students. The grand mean by gender for the full dataset. The whiskers represent a 95% exact confidence interval. The difference is significant ( $\chi^2$  test,  $p < .001$ ,  $N=5,239$ ).

## RESULTS

The data reveal large gender inequality in the probability to obtain a doctoral degree with a ‘cum laude’ distinction. Figure 1 shows that 6.57% of all male PhD students obtained cum laude, compared to 3.68% of all female students. Compared to dissertations of female PhD students, dissertations of male PhD students are 1.8 times ( $\chi^2$  test,  $p < .001$ ) more likely to be considered as belonging to the top 5%.

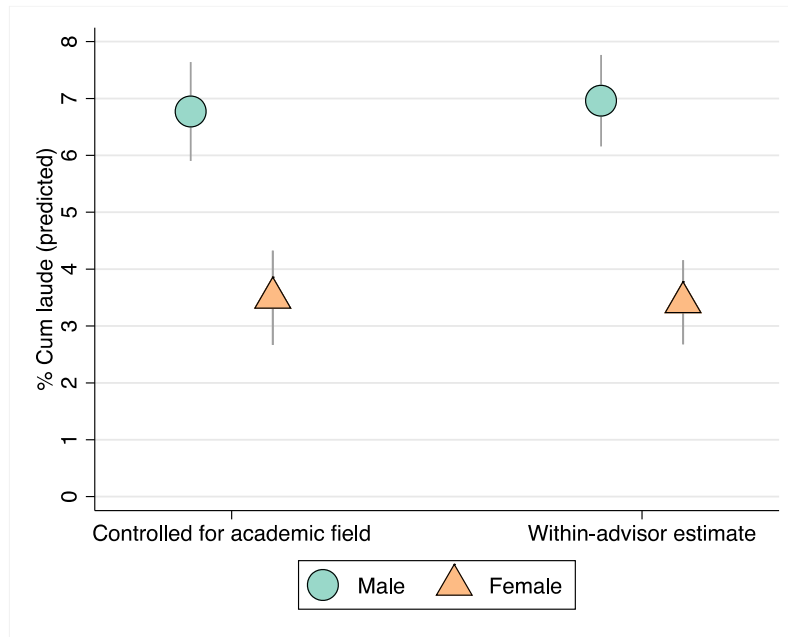
This large gender inequality in ‘cum laude’ distinctions does not necessarily support the idea that male and female PhD students are evaluated differently. Evaluation practices and definitions of quality vary strongly across scientific domains (23, 28). The observed inequality in Fig. 1 therefore could also be explained by a Simpson’s paradox (31): female scholars are more likely to obtain their PhD in academic fields where ‘cum laude’ distinctions are rarely given. Earlier research has for example shown that gender inequality in grant funding (32) or citations (33) can be explained by such a paradox.



**Fig. 2. ‘Cum laude’ distinctions across academic fields.** The relation between academic field and cum laude of PhD students. The left panel shows the relation between the percentage of cum laude distinctions (y-axis) and percentage of female PhD students (x-axis) for six academic fields (N=5,239). The size of the markers corresponds to the relative size of that field. The right panel depicts the gender gap in cum laude distinctions for each of the six fields (N=5,239).

Figure 2 evaluates whether differences across academic fields can explain the observed gender gap in cum laude distinctions. The left panel shows the association between the proportion of female PhD students and the likelihood to get a cum laude distinction across different academic fields. A systematic pattern is absent. In Medicine the probability to receive a ‘cum laude’ distinction is comparatively lowest (3.6%), while it is highest in the Social Sciences (9.2%). In both fields most PhD students are female. If anything, it seems that ‘cum laude’ is more often

awarded in academic fields with more female PhD students. If the gender inequality in Fig. 1 was explained by a Simpson’s paradox, we would expect the exact opposite. Figure 2 (right panel) shows that the gender gap in ‘cum laude’ is apparent across all six academic domains. In line with earlier studies (18), one might hypothesize that the female disadvantage in cum laude would be largest in the academic fields that include the least female academics (i.e., science, economics). While the magnitude of the gap differs, these cross-field differences are not statistically significant ( $p > .05$ ).



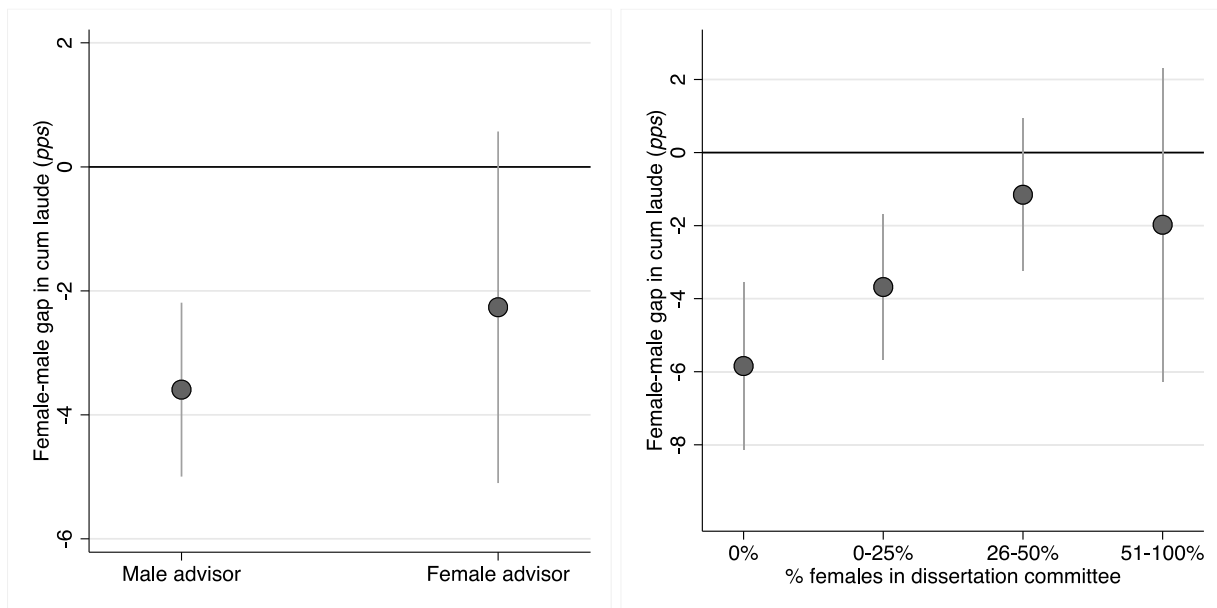
**Fig. 3. Estimates from regression models.** The right panel depicts the predicted gap from two regression analyses: (a) the predicted gap between male and female PhD students after taking academic field into account (‘Controlled for academic field’), and (b) the predicted gap from advisor fixed effects models. For each estimate a normal 95% confidence interval is displayed. The effect of ‘Controlled for academic field’ is significant ( $p < .001$ ,  $N=5,239$ ). The ‘Within-advisor estimate’ is significant too ( $p < .001$ ,  $N=7,249$ ). More information can be found in the *SM Appendix*, the underlying estimates can be found in Table S2 in the *SM Appendix*.

The absence of a Simpson’s paradox is confirmed by regression analyses (Fig. 3, left estimate) where the male advantage in ‘cum laude’ distinctions remains stable after controlling for academic field ( $p < .001$ ). Differences across academic fields do not explain the large gender disparity observed in Fig. 1.

It might be that the definition of fields used in Fig. 2 is too broad, for example, if there are selection effects into more narrow academic fields (33) or even doctoral advisors. If the supervisory quality of doctoral advisors or their likelihood to apply for a ‘cum laude’ correlates with the likelihood to supervise a female PhD student, this might explain the observed gender inequality. Figure 3 (right estimate) presents the results of a regression analysis that includes a fixed effect (30) for doctoral advisor (see *SM*). This estimate can be interpreted as the average within-advisor gender gap in ‘cum laude’ distinctions. It indicates whether male and female PhD

students with the same doctoral advisor had a different probability for a ‘cum laude’ distinction. In this fixed effects specification, the largest gender gap is unchanged ( $p < .001$ ). Dissertations of female PhD students are, on average, almost twice less likely to get the mark of excellence than their male counterparts with the same doctoral advisor.

5 To what extent do we observe gender-driven homophily in ‘cum laude’ distinctions? Figure 4 evaluates this by looking both at the gender of the doctoral supervisors (left panel) and the gender composition of the committee that evaluates the dissertation (right panel). In both figures the y-axis presents the female-male gap in cum laude distinctions in percentage points ( $pp$ ). For doctoral advisors (Fig. 3, left panel), we find little evidence for gender bias homophily. Female PhD students are less likely than male PhD students to obtain ‘cum laude’ irrespective of whether their doctoral advisor is male (3.6 $pp$  less likely) or female (2.3 $pp$  less likely). While the gender gap in cum laude distinctions is smaller when the advisor is female, this difference between male and female doctoral advisors is not statistically significant ( $p > .05$ ).



15 **Fig. 4. Differences in the gender gap across evaluators.** The association between respectively the gender of the doctoral advisors (left panel) and the gender composition of the dissertation committee and the size of the gender gap in cum laude distinctions. The y-axis presents the female-male gap in cum laude distinctions. It indicates how many percentage points ( $pp$ ) female PhD students are less likely to obtain a cum laude distinction. Estimates are obtained from regression analyses that included controls for academic field and year. For each estimate, a normal 95% confidence interval is displayed. The estimates in the left panel are not significantly different from each other ( $p > .05$ ,  $N=7,219$ ). Between the estimates in the right panel there are significant differences ( $p < .05$ ,  $N=7,219$ ). More information can be found in the *SM Appendix*, the underlying estimates can be found in Table S3 in the *SM Appendix*.

25 The dissertation committee plays a crucial role in awarding the ‘cum laude’ distinction too: they must agree that the dissertation belongs to the top 5%. For the committee, we find a clear pattern of gender homophily (Fig. 4, right panel). These results are again based on a fixed effects specification and exploit the variation in committee composition *within* the same doctoral advisor. There is no exogenous variation in committees (34), and doctoral advisors have an

important say in determining the composition of the committee. However, the fixed effects approach rules out that the observed pattern of gender homophily stems from variation across doctoral advisors: all effects are estimated within doctoral advisors.

The gender gap in ‘cum laude’ is largest when the dissertation is evaluated by an all-male committee: 5.8pp. In this scenario, 9.1% of male PhD students obtain a ‘cum laude’ distinction, compared to 3.3% of female PhD students with the same doctoral advisor—a difference by a factor of almost 3 (see *SM Appendix* Figure S5). The gender gap in ‘cum laude’ distinctions decreases as the gender composition shifts towards a greater number of female committee members. For dissertations that are evaluated by a committee where at least a quarter of the members is female, the gender gap in ‘cum laude’ distinctions disappears. More balanced dissertation committees were a minority. Over the observed period, 9.6% of all dissertations were evaluated by a committee where at least half of the members were female, whereas 27.8% were evaluated by an all-male committee (see *SM Appendix* Table S1).

We have performed several robustness checks. First, some studies have argued that gender inequality in evaluation is decreasing over time (35), additional analyses show that the found pattern of gender inequality in ‘cum laude’ distinctions remains stable over the observed period (2011-2021) (see *SM Appendix* Figure S8). Second, the seniority of committee members might explain the homophily results here, given that female committee members might be less likely to be senior. We do not find evidence for this, and the presented results are robust when we include committee members’ seniority (see *SM Appendix* Table S5). Finally, next to doctoral advisors, PhD students often have co-supervisors (see *SM Appendix*). They are involved in supervision but do not have the formal right to hand out doctoral degrees. Our finding that there are no homophily effects for doctoral advisors holds when we look at the gender composition of the full supervisory team (see *SM Appendix* Figure S9).

## DISCUSSION

We conclude that there is substantial inequality in the extent to which the work of early researchers is evaluated and perceived as excellent. Male PhD students are almost twice as likely than female PhD students to obtain a ‘cum laude’ distinction. The gender of the doctoral advisors is not significantly associated with the size of the gap, but the gender composition of the committee is: the average gender gap in ‘cum laude’ distinctions is largest for all-male committees and nears zero when committees are getting closer to gender parity.

A limitation of the current analysis is that we do not know which committee member has voted for or against ‘cum laude’, which means that we were only able to estimate homophily effects in the composition of the evaluation team. Moreover, we do not have any information on the two external anonymous reviewers that also have to attest that the dissertation deserves ‘cum laude’. Although the names of these external referees are provided by the doctoral advisor(s), we cannot rule out that they also contribute to the observed gender inequality in ‘cum laude’ distinctions. Earlier studies also pointed to the unequal effects of parenthood to understand gender inequality



in academia (36). While we do not have data on parenthood, only a small fraction of PhD students become parents during the writing of their dissertation (37, 38).

Most prominently, with the current data we are unable to determine whether the observed gender inequality in cum laude distinctions is explained by gender bias (39). It might for example be that male PhD students are overrepresented in the right tail of the distribution. Existing research provides little evidence for this. Among school-aged children, both boys and girls are overrepresented in the right tail depending on which skill is tested (numeracy or literacy) (40, 41). Moreover, since selection into PhD programs in the Netherlands mostly depends on grades and performance in college, it is highly implausible that male PhD students are of higher quality than female PhD students. To isolate the part of the effect that is driven by bias would be to “control” for the quality of the PhD dissertation (e.g., publications, citations for dissertational work). However, this will not help us further since these measures of quality are themselves affected by gender bias (21, 39, 42).

Another alternative explanation for our findings might be found in a gender gap in competitiveness (43, 44). However, PhD students are not aware that their doctoral advisors apply for ‘cum laude’ and have no influence over this decision. It is therefore highly unlikely that the gender gap in ‘cum laude’ distinctions, therefore, is driven by differences in strategic choices concerning the application process (45). Moreover, the reported gender gaps in competitiveness are much smaller than the effects found here.

While we are unable to interpret the observed gender inequality as bias, we do believe that our results provide strong indications for a biased perception of excellence. The gender gap in cum laude is not explained by differences across academic fields or doctoral advisors: within the same doctoral advisors, male PhD students are almost two times more likely to obtain a ‘cum laude’.

How big is the career advantage for those who obtain a ‘cum laude’ distinction? In other words, how important is it? While it is difficult to quantify this, graduating ‘cum laude’ is an official criterion of the largest Dutch grant funder to signal the quality of the researcher. More generally it is used as a mark of excellence, particularly in job applications for early career researchers. We believe that the findings of this study bear relevance besides the importance of ‘cum laude’, as the current study shows that already in the early career there is a large gender inequality in who is perceived as excellent. Men are more often deemed to belong to the top than women, particularly when evaluated by men.

Recent studies show that gender inequality in academia remains pervasive (21, 46). This study forms no exception. In line with earlier work (45), our research indicates that gender inequality is partly explained by the existence of institutional barriers that tend to be higher for women than for men. Obtaining a ‘cum laude’ distinction is such a barrier, that turns out to be easier to pass for men than for women. How to solve this inequality? The current study presents evidence for gender bias homophily in evaluators. A straightforward solution would therefore be to enhance the gender balance in evaluation committees in academia. However, given the gender segregation across academic fields (see Figure 2, left panel), for some academic disciplines this

will be unachievable in the short term—or it would put even greater pressure on the small group of female academics in those fields.

However, we believe that this study raises a more fundamental question: is it always crucial to distinguish the excellent from the rest? In some cases, such as hiring, it is unavoidable: there are not enough jobs for everyone. But academia is rife with prizes, awards, and distinctions, that often serve no other purpose than marking some as excellent. Designating five percent of all dissertations as ‘cum laude’ provides little functional benefits for academia, but it does perpetuate existing inequalities. Debates about bias in the evaluation of excellence therefore should not just be about how to create equal opportunities when men and women face institutional barriers in academia, but also about the necessity of institutional barriers in the first place.

## MATERIALS AND METHODS

### Data

In this study, we analyze information from one of the largest public universities in The Netherlands. The data contained information on all PhD students that completed their dissertations at the university between 2011 and 2021. In total, 5,240 dissertations were written over this period. All dissertations are part of the data, which means that we analyze population data for the university.

The university under study is one of the largest research universities in Europe. It employs over 3,000 researchers and is home to 40,000 students. The university is organized into seven faculties that cover the breadth of academic research: Medicine, Law, Social Sciences (e.g., Sociology, Psychology), Humanities (e.g., Philosophy, Language, History), Economics and Business, and Science (e.g., Physics, Mathematics, Biology). All large academic disciplines are represented in the university, with engineering as the exception. In the Netherlands, Engineering Schools can be found in the three technical universities, and hence PhD dissertations written in the technical domain were not part of the data.

The data consists of 5,239 dissertations—one observation is set to missing because no doctoral advisors were registered. PhD students can have multiple doctoral advisors, and many doctoral advisors have supervised multiple PhD students in the observed period (2011-2021) (see Figure S4 in the *SM Appendix*).

In our data, we identified 1,623 unique doctoral advisors  $j$ , that have supervised a total of 5,239 PhD students  $i$ . In total, there were 7,249 doctoral advisor–PhD student combinations  $ij$  in our data. In the regression analyses, we analyzed different samples, depending on the method or research question. For example, we analyzed the sample of PhD students ( $N=5,239$ ) when we were interested in variation between PhD students (for example in calculating the raw difference in ‘cum laude’ distinctions between men and women), but the sample of all PhD student–doctoral advisor combinations ( $N=7,249$ ) were analyzed when we were interested in the role of the doctoral advisors or committee.

We have collected different information on the dissertation and the PhD student from the university databases. First, for each dissertation we knew the gender of the PhD student, the date of obtaining the PhD, the broad academic field the dissertation was written in, and whether the dissertation was awarded cum laude. Second, for the supervisors, we knew their roles in the supervision (doctoral advisor or co-supervisor), names, gender, and academic title. This latter information allows us to separate full professors from assistant and associate professors, as only the first is allowed to carry the "Prof." title. The academic title thereby is a good proxy of the seniority of the doctoral advisor (or co-supervisor). Finally, the data contained information on the dissertation committee. For them, we have collected their gender and academic title. Table S1 in the *SM Appendix* provides an overview of all measures used in the study.

### Statistical analyses

Some results in the article are descriptive, but others are obtained by performing different types of regression analyses. In all analyses, we estimated a linear probability model (47, 48) on our binary dependent variable  $y_i$ , which measures whether the dissertation of PhD student  $i$  did (1) or did not (0) receive a 'cum laude' distinction. While we use linear probability models, logistic regressions provide the same findings (see *SM Appendix* Figure S10). To obtain the results presented in the article, we use three different regression models, depending on whether we analyze the sample of PhD students  $i$  or the sample of all combinations between PhD students  $i$  and doctoral advisors  $j$ .

First, we have used a simple linear regression at the level of PhD students ( $N=5,239$ ) to estimate the results in Figure 1 and the left estimates in Figure 3. Here we regressed obtaining cum laude on the gender of the PhD student. In this model, we could add several controls measured at the level of the PhD student (year of obtained PhD; academic field). Details on the model and the estimated results can be found in the *SM Appendix*.

Second, we have estimated multilevel regressions on the sample of all combinations between PhD students ( $i$ ) and doctoral advisors ( $j$ ) ( $N=7,239$ ). We used these models because we wanted to know whether gender inequality in 'cum laude' distinctions differed across male and female doctoral advisors (Figure 4, left panel). The data structure is cross-classified: doctoral advisors had multiple PhD students, but PhD students also had multiple doctoral advisors. To correct for this we cluster standard errors within PhD students. Details on the equation and the estimated results can be found in the *SM Appendix*

Finally, we used the combined sample of doctoral advisors ( $N=7,239$ ) to estimate fixed effects models. By including a fixed effect for the doctoral advisor  $j$ , all between-doctoral advisor variance is fixed, and the only variance that is left to explain was that within each advisor. This means that we estimated whether, on average, the same doctoral advisor was more likely to have male than female PhD students with 'cum laude' distinctions (Figure 3, right estimate). In a similar way, we exploited the variation in the gender composition of the dissertation committees within the same doctoral advisors and estimated whether, on average, for the same supervisors the gender inequality in 'cum laude' depended on the percentage of female members of the dissertation committee (Figure 4, right panel). An important benefit of these models is that we

were able to rule out the variation between doctoral advisors as a source of variation for gender inequality in 'cum laude'. More information on the fixed effects models and its assumptions, as well as the obtained results, can be found in the *SM Appendix*.

5 **Acknowledgments:** I would like to acknowledge H el ene Boeren for helping with the collection and harmonizing of the data, Amy Zeegelaar and Niek Brunsveld for discussions on potential explanations of gender inequality in 'cum laude' distinctions, and Sara Geven, Mathijs de Vaan, and Arnout van de Rijt for providing feedback to earlier drafts.

10 **Data and materials availability:** All code is available via [link to OSF repository]. The data will be made available in anonymized form. Full data access (including identifiable information) will only be possible when granted by the supplier of the university that was studied.

**Ethics:** The current study and data collection have been approved by the Ethical Committee of the Amsterdam Institute for Social Science Research of the University of Amsterdam.

15 **Supplementary Materials Appendix**

Description of 'cum laude' distinction

Data and methods

Results from regression results

Robustness checks

20 Table S1 – S5

Fig S1 – S10

References

## References and Notes

1. M. W. Nielsen, J. P. Andersen, Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences*. **118**, e2012208118 (2021).
2. B. F. Reskin, Scientific Productivity and the Reward Structure of Science. *American sociological review*, 491–504 (1977).
3. A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley, Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proceedings of the National Academy of Sciences*. **108**, 18–23 (2011).
4. T. Bol, M. De Vaan, A. Van de Rijt, The Matthew effect in science funding. *PNAS*. **115**, 4887–4890 (2018).
5. C. J. Lee, C. R. Sugimoto, G. Zhang, B. Cronin, Bias in peer review. *Journal of the American Society for Information Science and Technology*. **64**, 2–17 (2013).
6. C. Wenneras, A. Wold, Nepotism and Sexism in Peer-Review. *Nature*. **387** (1997).
7. S. Knobloch-Westerwick, C. J. Glynn, M. Huge, The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*. **35**, 603–625 (2013).
8. D. Murray, K. Siler, V. Larivière, W. M. Chan, A. M. Collings, J. Raymond, C. R. Sugimoto, Author-Reviewer Homophily in Peer Review (2019), p. 400515.
9. L. Bornmann, R. Mutz, H.-D. Daniel, Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*. **1**, 226–238 (2007).
10. T. Bol, M. de Vaan, A. van de Rijt, Gender-equal funding rates conceal unequal evaluations. *Research Policy*. **51**, 104399 (2022).
11. H. O. Witteman, M. Hendricks, S. Straus, C. Tannenbaum, Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet*. **393**, 531–540 (2019).
12. F. Mengel, J. Sauermann, U. Zölitz, Gender bias in teaching evaluations. *Journal of the European economic association*. **17**, 535–566 (2019).
13. D. Storage, Z. Horne, A. Cimpian, S.-J. Leslie, The frequency of “brilliant” and “genius” in teaching evaluations predicts the representation of women and African Americans across fields. *PloS one*. **11**, e0150194 (2016).
14. A. Boring, Gender biases in student evaluations of teaching. *Journal of Public Economics*. **145**, 27–41 (2017).
15. L. A. Rivera, A. Tilcsik, Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation. *Am Sociol Rev*. **84**, 248–274 (2019).
16. K. Dutt, D. L. Pfaff, A. F. Bernstein, J. S. Dillard, C. J. Block, Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience*. **9**, 805 (2016).
17. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, Science faculty’s subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*. **109**, 16474–16479 (2012).
18. F. Squazzoni, G. Bravo, M. Farjam, A. Marusic, B. Mehmani, M. Willis, A. Birukou, P. Dondio, F. Grimaldo, Peer review and gender bias: A study on 145 scholarly journals. *Science advances*. **7**, eabd0299 (2021).
19. C. W. Fox, C. T. Paine, Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. *Ecology and Evolution*. **9**, 3599–3619 (2019).
20. M. Lutter, M. Schröder, Who becomes a tenured professor, and why? Panel data evidence from German sociology, 1980–2013. *Research Policy*. **45**, 999–1013 (2016).

21. V. Larivière, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: Global gender disparities in science. *Nature*. **504**, 211–213 (2013).
22. H. Sarsons, K. Gërkhani, E. Reuben, A. Schram, Gender differences in recognition for group work. *Journal of Political Economy*. **129**, 101–147 (2021).
- 5 23. M. Lamont, *How professors think* (Harvard University Press, 2009).
24. J. M. Sheltzer, J. C. Smith, Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences*. **111**, 10107–10112 (2014).
25. M. Teplitskiy, D. Acuna, A. Elamrani-Raoult, K. Körding, J. Evans, The sociology of scientific validity: How professional networks shape judgement in peer review. *Research Policy*. **47**, 1825–1841 (2018).
- 10 26. A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Science advances*. **1**, e1400005 (2015).
27. European Commission, “She Figures: Gender in research and innovation” (European Commission, Brussels, 2013).
- 15 28. M. van den Brink, Y. Benschop, Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*. **19**, 507–524 (2012).
29. P. O’Connor, C. O’Hagan, Excellence in university academic staff evaluation: a problematic reality? *Studies in Higher Education*. **41**, 1943–1957 (2016).
30. P. D. Allison, *Fixed effects regression models* (SAGE publications, 2009), vol. 160.
- 20 31. P. J. Bickel, E. A. Hammel, J. W. O’Connell, Sex Bias in Graduate Admissions: Data from Berkeley: Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation. *Science*. **187**, 398–404 (1975).
32. C. J. Albers, Dutch research funding, gender bias, and Simpson’s paradox. *PNAS*. **112**, E6828–E6829 (2015).
- 25 33. A. Tekles, K. Auspurg, L. Bornmann, Same-gender citations do not indicate a substantial gender homophily bias. *PLOS ONE*. **17**, e0274810 (2022).
34. M. De Paola, V. Scoppa, Gender discrimination and evaluators’ gender: evidence from Italian academia. *Economica*. **82**, 162–188 (2015).
35. S. J. Ceci, W. M. Williams, Understanding current causes of women’s underrepresentation in science. *Proceedings of the National Academy of Sciences*. **108**, 3157–3162 (2011).
36. A. C. Morgan, S. F. Way, M. J. Hofer, D. B. Larremore, M. Galesic, A. Clauset, The unequal impact of parenthood in academia. *Science Advances*. **7**, eabd1996 (2021).
37. T. Bol, B. Derks, L. Poorthuis, “The impact of the COVID-19 pandemic first lockdown period on the work and well-being of academics in the Netherlands” (De Jonge Akademie/LNVH, Amsterdam/Utrecht, 2021).
- 35 38. Central Bureau of Statistics, “Resultaten Gepromoveerdenonderzoek 2019” (Report, The Hague, the Netherlands, 2020), (available at <https://www.cbs.nl/nl-nl/achtergrond/2020/14/resultaten-gepromoveerdenonderzoek-2019>).
39. V. A. Traag, L. Waltman, Causal foundations of bias, disparity and fairness (2022), , doi:10.48550/arXiv.2207.13665.
- 40 40. S. Machin, T. Pekkarinen, Global Sex Differences in Test Score Variability. *Science*. **322**, 1331–1332 (2008).
41. D. Reilly, D. L. Neumann, G. Andrews, Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*. **74**, 445 (2019).
- 45 42. S. F. Way, D. B. Larremore, A. Clauset, "Gender, productivity, and prestige in computer science faculty hiring networks" in *Proceedings of the 25th international conference on world wide web* (2016), pp. 1169–1179.

43. T. Buser, M. Niederle, H. Oosterbeek, Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics*. **129**, 1409–1447 (2014).
44. C. Bosquet, P.-P. Combes, C. García-Peñalosa, Gender and promotions: Evidence from academic economists in France. *The Scandinavian Journal of Economics*. **121**, 1020–1053 (2019).
- 5 45. A. L. Conner, K. S. Cook, S. J. Correll, H. R. Markus, C. A. Moss-Racusin, C. B. Muller, J. L. Raymond, C. Simard, Obscuring Gender Bias with “Choice.” *Science*. **343**, 1200–1200 (2014).
- 10 46. S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, A.-L. Barabási, Science of science. *Science*. **359**, eaao0185 (2018).
47. J. D. Angrist, J.-S. Pischke, *Mostly harmless econometrics: An empiricist’s companion* (Princeton university press, 2009).
- 15 48. P. D. Allison, R. A. Williams, V. Hippel, Better predicted probabilities from linear probability models. *Statistical Horizons* (2020).

Supplementary Materials Appendix for  
**Gender inequality in ‘cum laude’ distinctions for  
PhD students**

Thijs Bol



# Contents

- 1 Description of 'cum laude' application 3**
  - 1.1 Description of PhD trajectory in the Netherlands . . . . . 3
    - 1.1.1 Supervisors . . . . . 3
    - 1.1.2 Dissertation committee . . . . . 4
    - 1.1.3 Cum laude procedure . . . . . 6
  
- 2 Data and methods 7**
  - 2.1 Measures . . . . . 8
  - 2.2 Methods . . . . . 9
  
- 3 Results from regression models 12**
  - 3.1 Main effect of gender of PhD . . . . . 12
  - 3.2 Interaction analyses with advisors and dissertation committee . . . . . 13
  
- 4 Robustness checks 17**
  - 4.1 Trend over time . . . . . 17
  - 4.2 Role of co-supervisors . . . . . 18
  - 4.3 Seniority of committee . . . . . 19
  - 4.4 Logistic regressions . . . . . 21

# 1 Description of 'cum laude' application

## 1.1 Description of PhD trajectory in the Netherlands

In the Dutch system, PhD students are paid employees that get paid for three to four years to finish their dissertation. Depending on the academic field the PhD student is in, they can obtain a doctoral degree by writing journal articles or a monograph, where the latter is most common in the Humanities and some parts of the Social Sciences. PhD trajectories normally take 3 or 4 years, although delays do happen.

In the following we will describe the formal characteristics from the start of a PhD project to obtaining the doctoral degree, as laid out the doctoral regulations for the university under study (see Section 2 for more info on the specific case). Across universities in the Netherlands there is some variation in doctoral regulations, although the process is generally very similar to the one described here.

### 1.1.1 Supervisors

In the supervisory team of the PhD student supervisors can take two roles: the doctoral advisor, and the co-supervisor (also known as the "co-promotor"). All PhD students have at least one doctoral advisor, as they are the only ones with the *ius promovendi*, the legal right to act as main doctoral advisor (the "promotor") and award PhD students a doctoral degree. The doctoral board of the university will only allow a PhD student to enter the doctoral program when there is at least one doctoral advisor with the *ius promovendi*.

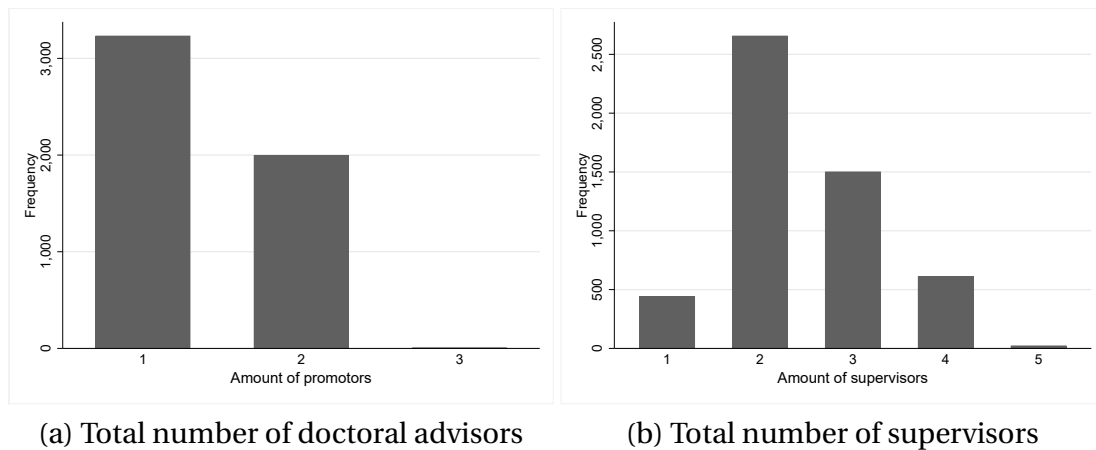
In the period under study, the right to act as doctoral advisor was only given to full professors.<sup>1</sup> PhD students can have multiple doctoral advisors, although it rarely happens that there are more than two doctoral advisors. Fig S1a shows that in our data the majority (62%) of PhD students had only one doctoral advisor. 38% had two doctoral advisors, and in 5 cases (0.1%) there were three full professors acting as doctoral advisor.

Besides their doctoral advisor, co-supervisors are often also part of the supervisory team of a PhD student. Co-supervisors are not obliged to have the *ius promovendi*, and they do not have to be a full professor. To the contrast, very often co-supervisors are assistant or associate professors. The only requirement that the doctoral board of the university has for co-supervisors is that they possess a doctorate themselves. In our data, the total supervisory team (doctoral advisor(s) + co-supervisor(s)) consisted of 2-4 people (see Figure S1b). 8% of all dissertations was supervised by only a doctoral

---

<sup>1</sup>In the past two years, associate professors can also get the *ius promovendi* when certain conditions that are specific to universities and/or faculties are met.

Figure S1: Supervision of PhD students



advisor. The Doctoral Board of the University has to approve of the full supervisory team, although this is a formality when the above-mentioned conditions are met.

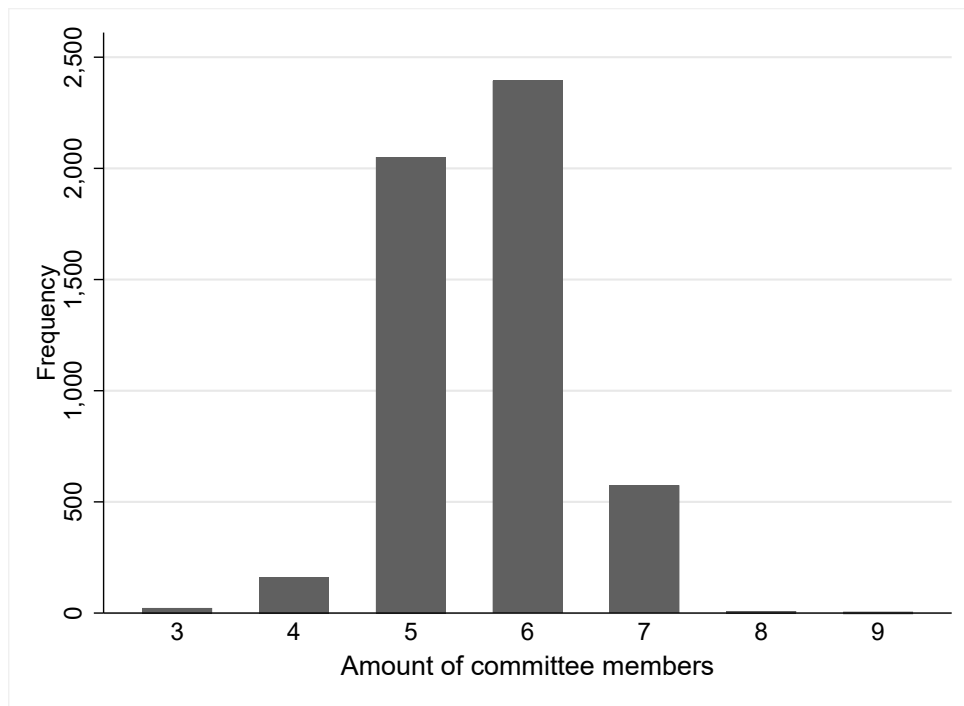
### 1.1.2 Dissertation committee

When the PhD thesis is almost finished, the supervisory team submits a proposal for the composition of the dissertation committee that will evaluate the dissertation and decide whether the candidate deserves the doctoral degree. The dissertation committee normally consists of at least five and at most seven members that all have voting rights about awarding the doctorate. Figure S2 shows that in our data the dissertation committees almost always consisted of 5-7 members, although a small fraction of the dissertation had committees with less than 5 members (3.5%) or more than 7 members (0.2%).

The supervisory team has the discretionary space to propose members of the dissertation committee, but there are some requirements. The majority of the voting members must be full professor, and at least half of them must be affiliated to the university where the doctoral degree is obtained. At least one member of the dissertation committee has to be a full professor working in the same faculty as where the PhD will get the doctorate, and at least one member must be from another (foreign) university. There are no rules with respect to the demographic composition (e.g., gender) of the dissertation committee.

When assessing the dissertation, committee members are asked to pay attention to several factors, including the clarity and relevance of the research puzzle, the robustness of the methods, the structure of the thesis and the extent to which the thesis is innovative. At least eight weeks before the defence ceremony, the committee members submit their vote for or against admitting the PhD student to the PhD ceremony. This vote cannot be conditional upon revisions in the dissertation. In contrast to many other countries, dissertation committees in the Netherlands face

Figure S2: Committee members



a binary decision: either they approve the dissertation or they do not. This means that there is no room for small adjustments or revisions, and the dissertation as handed in has to be approved or rejected. As an outcome, a vote against admitting the dissertation to the defence ceremony only happens when a committee member observes (a) major flaw(s) in the manuscript. Admission to the PhD ceremony is decided by a majority vote.

At the defence ceremony, the PhD student engages in a discussion with all members of the dissertation committee. The defence takes exactly 45 minutes, and in this time the committee members get the opportunity to ask questions and voice concerns, whereas the PhD student gets the opportunity to defend his/her/their thesis. The ceremony is public, and everyone interested is allowed to join.

After the defence, the committee retires to discuss the defence and dissertation. This deliberation takes place behind closed doors with only the doctoral advisor(s), co-supervisors, and dissertation committee members present.<sup>2</sup> First, the doctoral advisor(s) and co-supervisor(s) advise to the committee whether they believe the candidate should be awarded the doctorate and then all committee members vote for or against awarding the doctoral degree. This again is decided by a majority vote. Since the decision to allow the candidate to the ceremony has already been taken at this point, it is extremely rare that the doctoral degree will not be awarded at the ceremony. It is so rare that it has not happened at the university that we analyzed for

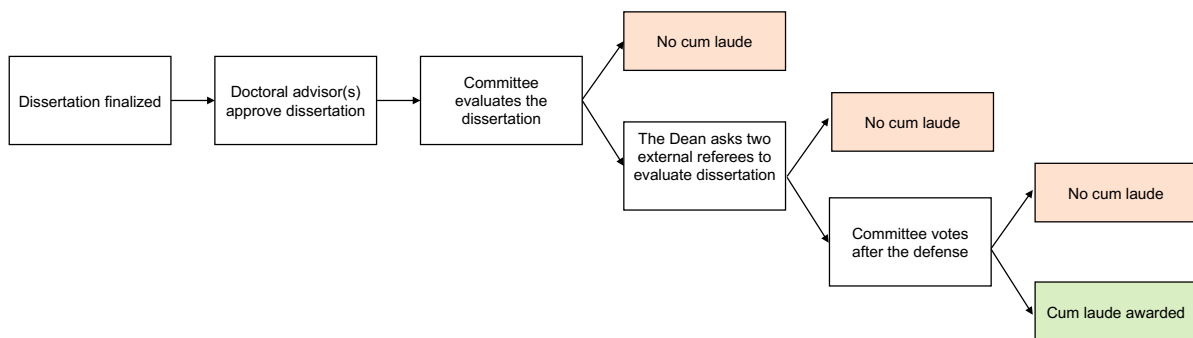
<sup>2</sup>There is also a chair of the promotion committee, but the chair has no voting rights and has presiding the defence and the deliberation afterwards as its main task.

the period under study (2011-2021). When the committee has made its decision, they return to the room of the public defense and award the PhD student with the doctoral degree.

### 1.1.3 Cum laude procedure

In this article we are mostly interested in 'cum laude' distinctions, but how does the awarding of a 'cum laude' works? In order to award dissertations there are several steps that need to be taken, and we will discuss all of them chronologically. A schematic overview of the full procedure can be found in Figure S3.

Figure S3: Schematic overview of cum laude procedure



After finalizing the dissertation, and the doctoral advisors have approved of the dissertation, it is send out to the dissertation committee, who will review the dissertation. After their evaluation of the dissertation, each committee member has to answer two questions related to 'cum laude': (a) whether they believe that this doctoral thesis should qualify for the distinction 'cum laude?', and (b) whether they can attest that the dissertation belongs to the top 5% of all dissertations that this committee member read in the relevant field of study. Both questions can only be answered with a simple yes or no.

When the assessments of the committee members are in, the doctoral advisor(s) can propose a 'cum laude' to the Dean of their Faculty. Such a 'cum laude' procedure can only be started when no more than one committee member answered "no" to (one of) the aforementioned two questions. If this criteria is met, the doctoral advisor has to write a detailed and reasoned proposal to the Dean explaining why 'cum laude' is merited. In that proposal, the doctoral advisor has to name at least four independent external referees. These referees cannot be member of the dissertation committee and the list of external referees has to include scholars working at foreign universities.

The Dean will, without consulting the doctoral advisor(s), choose two of the external referees from the list. These external referees remain anonymous to the doctoral advisor(s) and the PhD student. The two external referees read the dissertation and submit a motivated recommendation explaining whether or not the dissertation de-

serves 'cum laude'. Both external referees have to agree that the dissertation belongs to the top 5% of that respective field: if one external referee gives a negative advice, the 'cum laude' is off the table.

If both external referees give a positive advice, the dissertation committee has the deciding vote. However, only after the PhD defence they will vote whether the dissertation receives 'cum laude'. This process takes place behind closed doors, at the moment that the dissertation is deliberated by the doctoral advisor(s), co-supervisor(s), and committee members. Before the defence starts, the university prepares two different doctoral degrees: one with, and one without 'cum laude' on it. After the defence, the doctoral advisor(s) and co-supervisor(s) advice the committee on whether they believe the dissertation deserves 'cum laude'. The dissertation committee members then anonymously vote in writing. After all votes are in, they are counted by the chair of the dissertation committee. When no more than one vote is against, the PhD student receives the doctoral degree with 'cum laude' on it. The other doctoral degree, without 'cum laude' on it, is torn apart and thrown away.

When the discussion behind closed doors is ended, the dissertation committee and supervisors return to the public room of the defence where the PhD student gets awarded his or her doctoral degree. It is only at this point that the PhD student will learn whether the dissertation receives 'cum laude'. Since the final decision for 'cum laude' will only be made just before awarding the doctoral degree, PhD students are not being made aware of all the steps that are to taken before. PhD candidates do not know if 'cum laude' is on the table at the start of the defence, but only will learn afterwards.

## 2 Data and methods

The data consisted of 5,239 dissertations.<sup>3</sup> PhD students can have multiple doctoral advisors, and a large number of doctoral advisors have supervised multiple PhD students in the observed period (2011-2021) (see Figure S5).

Figure S4: Schematic overview of data

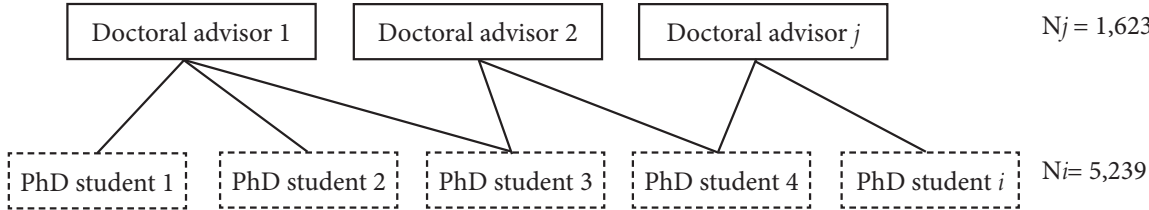


Figure S4 provides a schematic overview of the data. In our data we identified

<sup>3</sup>One observation is set to missing because no doctoral advisors were registered.

1,623 unique doctoral advisors  $j$ , that have supervised a total of 5,239 PhD students  $i$ . In total there were 7,249 doctoral advisor–PhD student combinations  $ij$  in our data. This means that in the observed period a PhD student had on average 1.38 doctoral advisors. In the regression analyses (see below) we have analyzed different samples, depending on the method or research question. For example, we analyzed the sample of PhD students ( $N=5,239$ ) when we were interested in variation between PhD students (for example in calculating the raw difference in 'cum laude' distinctions between men and women), but the sample of all PhD student–doctoral advisor combinations ( $N=7,249$ ) were analyzed when we were interested in the role of the doctoral advisors or committee. For each analysis used in the article we will specify the sample it was estimated on in Section 3.

## 2.1 Measures

Table S1 provides an overview of all measures used in the study. It is important to note that the sample size is not fully constant across all measures. This is mostly due to the level at which the variables are measured (the level of the doctoral advisor or the level of the PhD student). There are 24 missing observations (0.5%) for the gender composition of the dissertation committee, as for these dissertations the committee was not registered.

Table S1 makes clear that overall, 5% of all dissertations received 'cum laude', and more than half of the PhD students were female. Dissertation committees rarely contained more than 50% females: only in 10% of the dissertations, the majority of the dissertation committee was female. 28% of all PhD students defended their thesis in front of a committee consisting solely of men. The table also makes clear that the amount of PhD dissertations that were written remains quite stable over the observed period. The exceptions here were 2011 and 2021, but this had to do with data availability: data was not available for the full calendar years at the moment of data collection.

Finally, it is important to note that we do not know for each individual committee member whom voted for or against 'cum laude'. As explained above, the voting after the PhD defence takes place anonymously, which makes it impossible to track whether someone voted against. We do know that this can never be more than one committee member, as otherwise the 'cum laude' would not be awarded. Finally, in the process of awarding a 'cum laude', two external referees must also agree that the dissertation belongs to the top 5% of that field. The university does not systematically register the reports of these external referees, and therefore they were not part of the analyses.

In our data, only one in five of the doctoral advisors was female. Given that the *ius promovendi* was only given to full professors in the observed period, the low

Table S1: Description of data

	Mean	Range	SD	N
<b>PhD student</b>				
Cum laude (1=yes)	.05	0-1		5,239
Gender PhD (1=female)	.52	0-1		5,239
Academic domain				
Economics and Business	.04	0-1		5,239
Humanities	.10	0-1		5,239
Law	.03	0-1		5,239
Medicine	.45	0-1		5,239
Social Sciences	.15	0-1		5,239
Science	.22	0-1		5,239
# Doctoral advisors	1.38	1-3	.49	5,239
# Total supervisors	2.45	1-5	.82	5,239
Dissertation committee				
0% Female	.28	0-1		5,215
1-25% Female	.32	0-1		5,215
26-50% Female	.31	0-1		5,215
51-100% Female	.10	0-1		5,215
Year of defense				
2011	.07	0-1		5,239
2012	.09	0-1		5,239
2013	.10	0-1		5,239
2014	.09	0-1		5,239
2015	.09	0-1		5,239
2016	.10	0-1		5,239
2017	.10	0-1		5,239
2018	.10	0-1		5,239
2019	.10	0-1		5,239
2020	.10	0-1		5,239
2021	.04	0-1		5,239
<b>Doctoral advisor</b>				
Gender (1=female)	.19	0-1		1,628
# of PhD students	4.45	1-58	5.72	1,628

percentage of female doctoral advisors is most likely explained by the low number of female full professors in the Netherlands [1]. Finally, Table S1 shows that, on average, doctoral advisors have supervised multiple PhD students in the observed time period. On average, they have supervised 4.45 PhD students, with a maximum of 58. This variation is important for the fixed effects models, as explained below.

## 2.2 Methods

We use different regression models, depending on whether we analyze the sample of PhD students  $i$  or the sample of all combinations between PhD students  $i$  and doctoral advisors  $j$ .

The first model is estimated with a simple linear regression where our binary dependent variable  $y_i$ , which measures whether the dissertation of PhD student  $i$  did (1) or did not (0) receive a 'cum laude' distinction. The basic model in Equation 1



shows that  $y_i$  was regressed on  $X_{1i}$ , a dummy indicator measuring the gender of the PhD student. The key parameter of interest here was  $\beta_1$ , which provides the predicted effect of gender of the PhD student on obtaining 'cum laude'. The vector  $\zeta_i$  denotes the effects for different control variables that are added (see Table S2 for all controls that are added),  $\epsilon_i$  denotes the residual term.

$$y_i = \alpha_i + \beta_1 X_{1i} + \zeta_i + \epsilon_i \quad (1)$$

Second, in order to investigate whether gender inequality in 'cum laude' distinctions differed across male and female doctoral advisors, we estimated multilevel models. Here we used the sample of all combinations between PhD students ( $i$ ) and doctoral advisors ( $j$ ). Equation 2 depicts the statistical model that was used to answer this question. The key difference with Equation 1 is the presence of term  $\beta_2$ , that denotes the predicted effect of the gender of the doctoral advisor  $X_{2j}$ , and the interaction term  $\beta_3$  between the gender of the doctoral advisor and the gender of the PhD student. The vectors  $\zeta_{ij}$  and  $\rho_j$  include control variables measured at the level of the PhD student or the doctoral advisor. The data structure is cross-classified: doctoral advisors had multiple PhD students, but PhD students also had multiple doctoral advisors. To correct for this we cluster standard errors within PhD students, to get the most robust estimate of the predicted effect.

$$y_{ij} = \alpha_{ij} + \beta_1 X_{1ij} + \beta_2 X_{2j} + \beta_3 X_{1ij} * X_{2j} + \zeta_{ij} + \epsilon_{ij} \quad (2)$$

Third, we estimated fixed effects models. By including a fixed effect for the doctoral advisor  $j$ , all between-doctoral advisor variance is fixed, and the only variance that is left to explain was that within each advisor. Equation 3 presents the basic setup of this fixed effects model. It is similar to Equation 1, but has as the key exception the term  $\delta_j$ , which depicts a fixed effect for doctoral advisor.

$$y_{ij} = \alpha_{ij} + \beta_1 X_{1ij} + \zeta_{ij} + \delta_j + \epsilon_{ij} \quad (3)$$

An important benefit of these fixed effects models is that we were able to rule out variation between doctoral advisors as a source of variation for gender inequality in 'cum laude'. In order to be able to estimate such a fixed effects models, it is important to have sufficient variation within doctoral advisors. In other words: the data had to contain enough doctoral advisors that supervised multiple PhD students. Figure S5 shows the frequency of doctoral advisors depending on the number of completed dissertations in the observed period (2011-2021). A bit over 40% of all doctoral advisors have only supervised one finished dissertation in the observed period. For these doctoral advisors there is no within-variance, which means that they did not contribute to the estimated effect of  $\beta_{1ij}$  in Equation 3. At the same time, almost 60% of the

doctoral advisors did supervise multiple PhD students.

Figure S5: PhD students per doctoral advisors

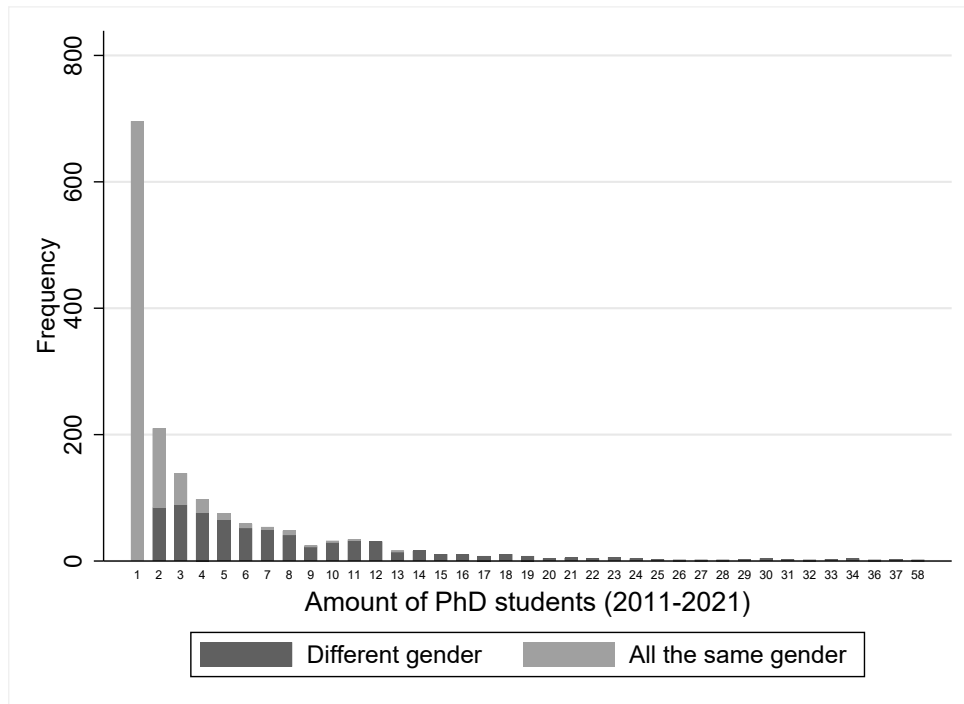


Figure S5 also shows whether the PhD students within one doctoral advisor had the same gender. Per definition, all doctoral advisors that finished supervising just one PhD dissertation belong to the group of doctoral advisors without gender variation among their supervisees. As the number of PhD students per supervisor increase, the likelihood of them all having the same gender decreases. Of the 1,628 doctoral advisors in the data, 43% has supervised multiple PhD students that do not all have the same gender. It is this group over which we are able to estimate the fixed effects models.

A worry might be that this group is selective, as strong bias might already play out in the hiring of PhD students. For that reason we also estimated the same models without including the fixed effect for doctoral advisor. As becomes clear from Tables S2 and S3, the estimates are very stable between the models with and without the fixed effects for doctoral advisors, indicating that the findings do not depend on the sample or method that was used.

### 3 Results from regression models

The estimates that are presented in Figure 1 and 2 in the article are based on simple descriptive statistics, whereas the estimates visualized in Figures 3 and 4 are derived from regression models. In this section we will discuss the outcomes of these regression models and some alternative specifications.

Table S2: Regression analyses on obtaining a 'cum laude' distinction

	(1)	(2)	(3)
Female PhD student	-0.029** (-4.78)	-0.033** (-5.26)	-0.035** (-5.82)
Academic field (ref= Econ. & Bus.)			
Humanities		0.021 (1.19)	0.170 (1.88)
Law		0.045* (2.01)	0.215* (2.30)
Medicine		0.005 (0.30)	0.138 (1.55)
Social Sciences		0.061** (3.64)	0.165 (1.85)
Science		0.006 (0.37)	0.084 (1.03)
Constant	0.064** (5.63)	0.049** (2.69)	-0.053 (-0.67)
Year dummies	Yes	Yes	Yes
Advisor FE	No	No	Yes
N	5,239	5,239	7,249

*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$

#### 3.1 Main effect of gender of PhD

Table S2 presents the estimates of the main effect of the the gender of PhD students on their probability to obtain 'cum laude'. In the linear probability model the coefficients can be interpreted as a marginal change in percentage points ( $pp$ ). Model 1 only included the gender of the PhD student and year dummies, to take out any general trend in gender inequality in 'cum laude' (see Section 4.1. Here we found that there is a baseline difference of 2.9 percentage points (-0.029) to the disadvantage of female PhD students.

In Model 2 we tested the Simpsons paradox (Figure 3), by including dummy variables for academic field. As visualized in the article, the gender gap in 'cum laude' distinctions is not affected by the inclusion by field of study and even becomes slightly

larger (3.3 *pp*). We did find some differences in the likelihood for 'cum laude' across academic fields. It is significantly larger in Law and Social Sciences compared to the reference category (Economics and Business), something that also descriptive became clear from Figure 2 (left panel).

Model 3, finally, exploited the within-doctoral advisor variation discussed in Equation 3. Here we included a fixed effect for doctoral advisor, and the predicted effect can therefore be interpreted as the average gender gap within doctoral advisors. After including advisor fixed effects, the estimate of the gender gap in 'cum laude' distinctions remains stable at -0.035 (3.5 *pp*).

### 3.2 Interaction analyses with advisors and dissertation committee

The next set of analyses (Table S3) dealt with the extent to which the gender gap in 'cum laude' depended on the doctoral advisor and the dissertation committee, as visualized in Figure 4 in the main article. Models 1 and 2 in Table S3 show the results of a regression where we looked at the extent to which the gender gap in 'cum laude' distinctions differs across male and female doctoral advisors. Model 1 makes clear that there was no main effect of having a female doctoral advisor: the probability of receiving 'cum laude' was not significantly associated with the gender of the doctoral advisor, and the effect was close to zero (0.005, or 0.5 *pp*). In Model 2 we added the interaction between the gender of the PhD student and the gender of the doctoral advisor. While the interaction effect was positive (0.013, or 1.3 *pp*), it does not reach conventional levels of statistical significance ( $p > 0.05$ ). As presented in Figure 4, the regression models provided no evidence for an association between the gender gap in 'cum laude' and the gender of the doctoral advisor.

Models 3-5 of Table S3 dealt with the role of the dissertation committee. Here we also added committee size as control variable, since it might be that doctoral advisors compose a relatively large committee when they believe the dissertation to be of excellent quality. Model 3 presents the main effects of the committee and the committee size. There are some differences in the likelihood of receiving 'cum laude'; the average probability to receive 'cum laude' was highest in all-male committees (the reference category). The likelihood of receiving a 'cum laude' was 1.7 *pp* (-0.017) smaller in a committee with up to 25% female members. There also was a positive association between committee size and 'cum laude': for each extra committee member the regression model predicted a 1.7 *pp* increase in the probability of receiving a 'cum laude' distinction. Since this model includes advisor fixed effects, it was an effect estimated within the same doctoral advisor.

Model 4 is presents the analyses that are visualized in the main article (Figure 4,

Table S3: Interaction analyses with advisor and committee

	(1)	(2)	(3)	(4)	(5)
Female PhD student	-0.034** (-4.93)	-0.036** (-5.02)	-0.033** (-5.37)	-0.058** (-5.00)	-0.059** (-5.47)
Female doctoral advisor	0.005 (0.75)	-0.003 (-0.26)			
Female PhD * Female advisor		0.013 (0.88)			
Committee composition (ref. 0% Female)					
1-25%			-0.017* (-2.17)	-0.026* (-2.49)	-0.026* (-2.15)
26-50%			-0.015 (-1.69)	-0.039** (-3.25)	-0.030* (-2.15)
51-100%			-0.025 (-1.94)	-0.046* (-2.10)	-0.047* (-2.00)
Committee size			0.017** (3.32)	0.017** (3.31)	0.015** (3.49)
Interaction w/ committee (ref. 0% Female)					
Female PhD * 1-25% Female committee				0.022 (1.42)	0.029* (2.01)
Female PhD * 26-50% Female committee				0.047** (2.99)	0.045** (2.78)
Female PhD * 51-100% Female committee				0.039 (1.56)	0.056* (2.13)
Constant	0.046** (2.60)	0.047** (2.62)	-0.133 (-1.56)	-0.119 (-1.40)	-0.017 (-0.60)
Year dummies	Yes	Yes	Yes	Yes	Yes
Academic field dummies	Yes	Yes	Yes	Yes	Yes
Clustered SEs	Yes	Yes	No	No	Yes
Advisor FE	No	No	Yes	Yes	No
N	7,249	7,249	7,219	7,219	5,215

*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$

right panel). It is the fixed effects model with the interaction between committee composition and gender of the PhD student. Model 4 makes clear that there was a significant interaction between the gender composition of the committee and the gender gap in 'cum laude' distinctions. The gap was largest (0.058 or 5.8*pp*) when the dissertation was evaluated by an all male committee. Compared to the 5.8, the gap decreases by 2.2*pp* when a committee includes up to 25% females, whereas the predicted gender gap is reduced by respectively 4.7*pp* and 3.9*pp* when the committee has 26-50% or 51-100% female members. In these committees the gender gap still exists, but becomes much smaller. Model 4 shows that the difference to the reference category (0% female members) is only significant when 26-50% of the committee members are female, the marginal predicted effects (main effect + interaction effect) that are portrayed in Figure 4 (left panel) tell a different story. Here it becomes clear that whenever a committee consists of at least 25% females, the gender gap in 'cum laude' is not significantly different from 0.

Figure S6: Predicted probability of 'cum laude' by committee composition

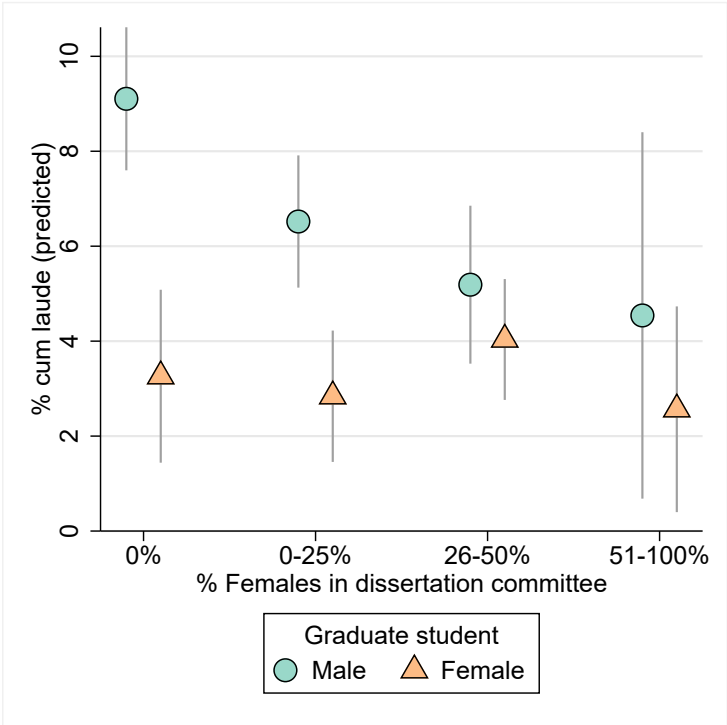


Figure S6 presents the same results as Figure 4 (right panel) as predicted probabilities for male and female PhD by the gender composition of the committee. Just like Figure 4 from main article, these results were based on Model 4 from Table S3. The figure makes clear that the larger gender gap in all-male committees was driven by a much higher share of male PhD students that received a 'cum laude' distinction. In all-male committees, a predicted 9.1% of male PhD students received cum laude, compared to 3.2% of female candidates—a difference by almost a factor 3. In line

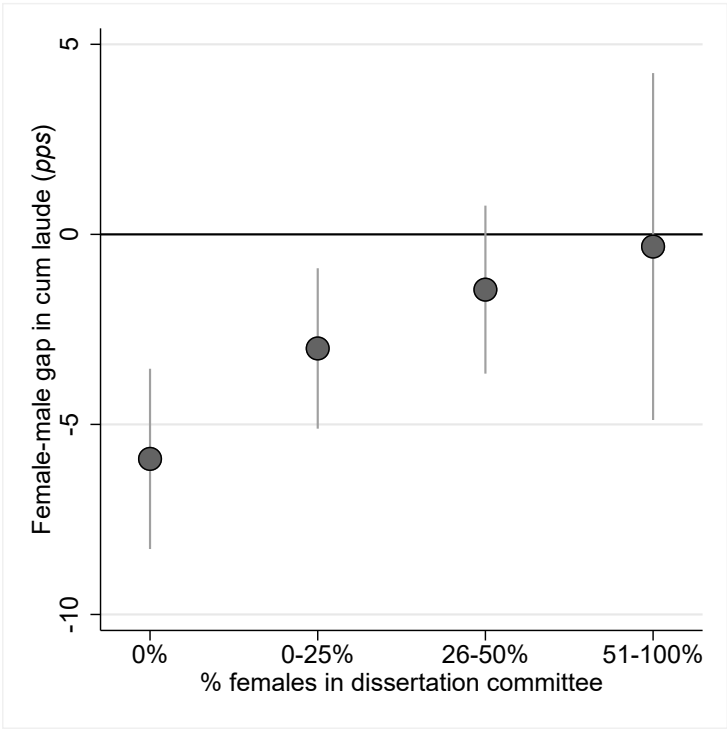
with Figure 4 from the article, the gap decreases and becomes insignificant as the committee got closer to gender parity.

The estimates in Models 3 and 4 were obtained from a fixed effects model. The model has large advantages, as it controls for all unobserved variation between doctoral advisors. However, a disadvantage is that the results are based on a selective subgroup of doctoral advisors, namely those who had multiple PhD students in the observed period. Moreover, one could argue that part of the gender gap in 'cum laude' could be driven by bias in selection of PhD students, for example when male doctoral advisors are more likely to hire male PhD students.

For this reason, Model 5 investigated the effect of the committee composition, by using the sample of dissertations (N=5,215). In this model, both the within- and between-advisor effects were captured. The estimates in Model 5 are remarkably similar to those in Model 4. If anything, the results more strongly confirms an association between the gender composition of the dissertation committee and the gender gap in 'cum laude': the gender gap in committees where at least 50% of the members is female disappeared completely.

In a similar way to Figure 4 (left panel), the marginal effects obtained from Model 5 are displayed in Figure S7. This figure shows an even stronger association between committee composition and the gender gap in 'cum laude', supporting the conclusions drawn in the main article. Moreover, it indicates that the findings presented in the paper are robust to the method that was used.

Figure S7: Gender gap in 'cum laude' by committee (Model 5, Table S3)



## 4 Robustness checks

### 4.1 Trend over time

In the article the gender gap in 'cum laude' distinctions was treated as static, even though the period under study spans over a decade (2011-2021). Earlier studies have argued that across several domains, gender inequality in science is decreasing, and gender bias is not as important of an explanation as it was some decades ago [4]. In order to investigate whether this argument also holds or our case, we have looked at whether the gender gap in 'cum laude' distinctions changed over time.

Table S4: Gender inequality in 'cum laude' over time

	(1)
Female PhD student	-0.040** (-3.33)
Year (0=2011)	-0.001 (-0.91)
Female * Year	0.001 (0.70)
Constant	0.057** (3.52)
Academic field dummies	Yes
N	5,239

*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$

Table S4 presents the results of a regression analysis including an interaction between gender of the PhD student and a continuous variable for year (starting at 2011). The table makes clear that there was main effect of year, indicating that the probability of obtaining a 'cum laude' distinction had no linear association to the year of obtaining the doctoral degree. More importantly, the interaction effect of year of completion and female on obtaining 'cum laude' was not significant ( $p > 0.05$ ) and not substantive in size: the predicted gender gap increased by 0.001 (or 0.1%) for every year.

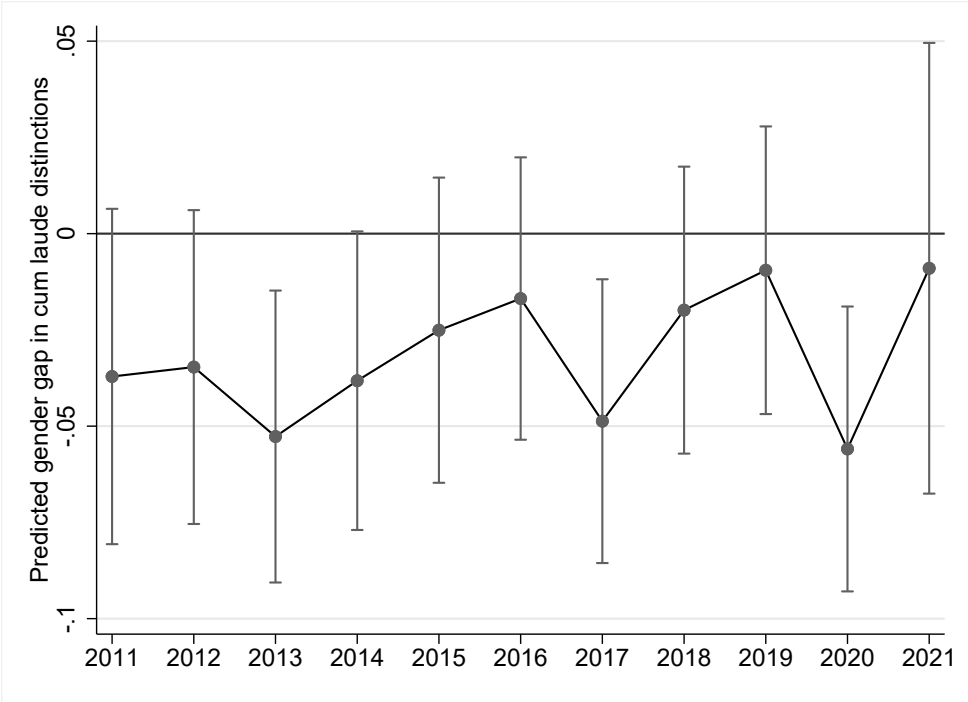
It might be the case that the interaction with year is not linear but does change over time. To model this we interacted the gender of the PhD student with a set of dummy variables for year.<sup>4</sup> Figure S8 presents the trend over time, where the estimates were obtained from a regression where year (measured as a set of dummy variables) was interacted with the binary indicator for gender (1=female). The figure makes clear

<sup>4</sup>This is also the way in which year was controlled for in the other analyses: as a set of dummy variables to allow for non linear effects.



that there were some differences between the years in the observed period, but they do not follow a clear pattern, nor does it seem to be the case that the gender gap in 'cum laude' has decreased in the decade under study. In the regression model there are no significant differences between any of the year combinations ( $p > 0.05$ ). In the case of 'cum laude' distinctions, we do not find evidence that it has decreased or even changed over time.

Figure S8: Gender gap in 'cum laude' by year



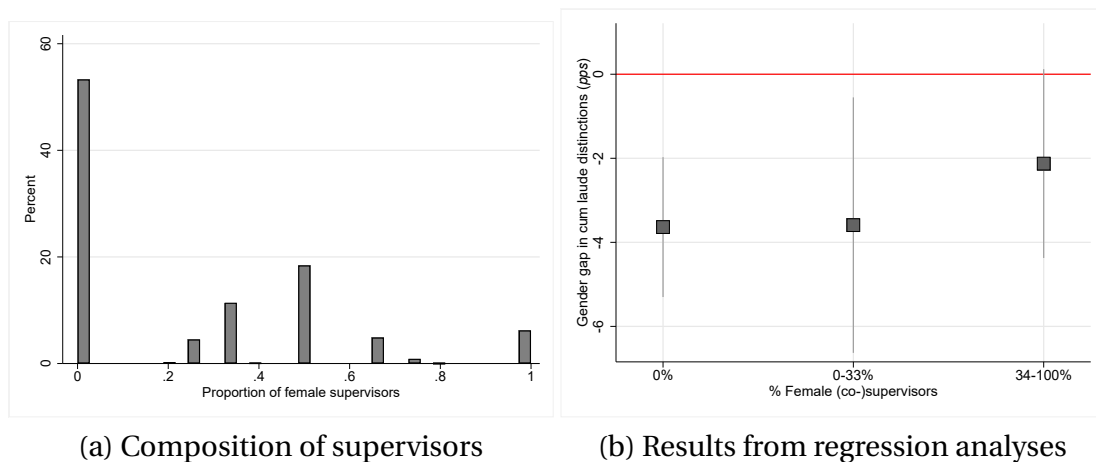
### 4.2 Role of co-supervisors

In Figure 4 (left panel) we have shown that the gender gap in 'cum laude' does not vary significantly between male and female doctoral advisors. As became clear from Section 1.1.1, almost all PhD students have co-supervisors next to their doctoral advisors. The co-supervisors were more likely to be female than the doctoral advisors: whereas only 19% of the doctoral advisors was female, 34% of all co-supervisors was female. It might be that the composition of the supervisory team does play a role, as here there will be more diversity than among the doctoral advisors.

For this reason we have re-estimated the analyses using the percentage of all supervisors that was female. For example, if someone had a male doctoral advisor, and two female co-supervisors, this percentage would be 67%. Figure S9a shows that the majority of supervisory teams consisted of only men: 53% of all PhD students had no female member in the supervisory team. Only 12% of the PhD students in the sample had a supervisory team where a majority of members was female.

Figure S9b shows whether there is an association between the composition of the supervisory team and the predicted gender gap in 'cum laude' distinctions. This figure is based on a regression model where a categorical variable of the gender composition of the supervisors (in three categories) was interacted with the gender of the PhD student (1=female). In line with what was found in Figure 4 (left panel), the effects are not significantly different across the three categories. The effect sizes are almost exactly the same when the supervisory team only consists of men or has at most 33% of female members. When more than one third of the members is female, the gender gap was smaller (about 1.7 times as small), but this difference is not significant ( $p > 0.05$ ).

Figure S9: Composition of supervisors



### 4.3 Seniority of committee

It might be that the likelihood to award a 'cum laude' distinction depends on the seniority of those who evaluate the thesis. One explicit question in the 'cum laude' evaluation is whether the thesis belongs to the best theses that evaluator has read in a given field. It could be the case that this probability increases with seniority: as scholars have evaluated more theses, they are better able to judge whether the given thesis is "excellent". Given that women in committee are less likely to be senior, given that many more men are full professor in the Netherlands [1], it might confound the association found between the proportion of female committee members and the gender gap in 'cum laude'.

In order to investigate this, we looked at the composition of the dissertation committee, and more specifically at what proportion of the committee members was full professor at time of evaluation. When the smaller gender gap in 'cum laude' in committee with more female members is confounded by their seniority, the current measure should provide an indication. Whether a given committee member is full professor or not is not registered for the full sample. Only from 2015 onward, this

information is available. This means that the analyses to answer this question were done on a smaller sample.

Table S5: Interaction analyses with advisor and committee

	(1)	(2)	(3)	(4)
Female PhD student	-0.057** (-3.47)	0.038 (0.77)	-0.064** (-3.91)	0.075 (1.54)
Committee composition (ref. 0% Female)				
1-25%	-0.023 (-1.73)	-0.022 (-1.62)	-0.034* (-2.51)	-0.032* (-2.33)
26-50%	-0.021 (-1.44)	-0.019 (-1.24)	-0.039* (-2.57)	-0.034* (-2.24)
51-100%	-0.019 (-0.75)	-0.014 (-0.52)	-0.031 (-1.18)	-0.021 (-0.80)
Interaction w/ committee (ref. 0% Female)				
Female PhD * 1-25% Female committee	0.032 (1.53)	0.031 (1.48)	0.036 (1.73)	0.034 (1.67)
Female PhD * 26-50% Female committee	0.044* (2.12)	0.041* (1.98)	0.053* (2.57)	0.049* (2.37)
Female PhD * 51-100% Female committee	0.033 (1.05)	0.026 (0.85)	0.031 (1.03)	0.020 (0.67)
Committee size	0.016** (2.65)	0.016** (2.78)	0.020** (2.93)	0.020** (2.98)
Professor composition committee		0.105* (2.56)		0.146** (3.53)
Female PhD student * Professor composition		-0.111* (-2.04)		-0.162** (-3.02)
Constant	-0.037 (-0.94)	-0.133* (-2.43)	-0.039 (-0.27)	-0.163 (-1.08)
Year dummies	Yes	Yes	Yes	Yes
Academic field dummies	Yes	Yes	Yes	Yes
Advisor FE	No	No	Yes	Yes
N	3,384	3,384	4,817	4,817

*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$

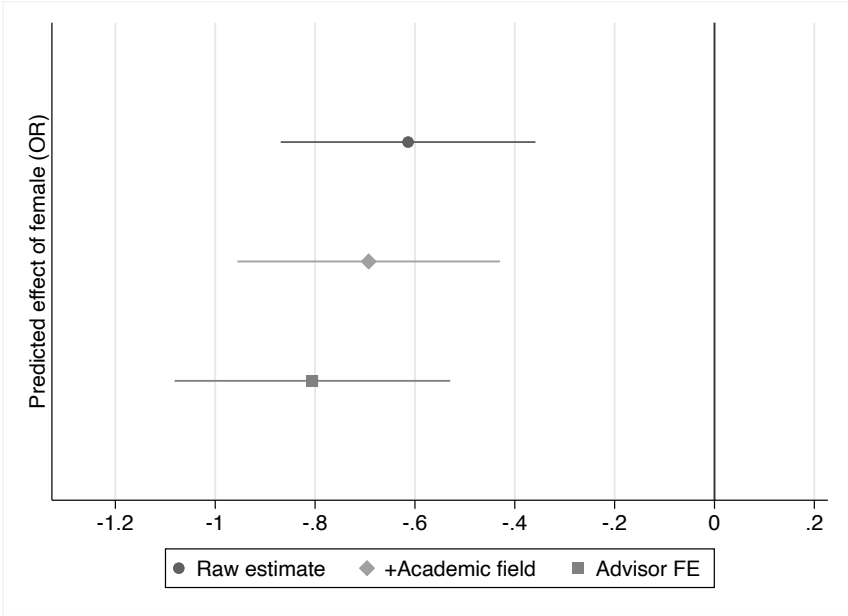
Table S5 shows the results of these analyses. There are two different models again: models on the sample of students  $i$  (Models 1 and 2), and fixed effects models (Models 3 and 4). From Model 2 it becomes clear that the association between the gender gap in 'cum laude' and the gender composition of the dissertation committee is not affected by the inclusion of the interaction with the professor composition of the committee. The point estimates (0.032, 0.044, 0.033) remain largely similar, and the significance does not change. There is an independent (interaction) effect of the professor composition in the dissertation committee. Model 2 shows that a male PhD

student is more likely to obtain 'cum laude' in a committee with a higher proportion of full professors. These findings are then replicated in Models 3 and 4, which use a fixed effects approach. In conclusion, the seniority of the committee is associated with the likelihood of female PhD students to obtain a 'cum laude' distinction, but this does not affect the main finding about the role of the female committee.

### 4.4 Logistic regressions

A final robustness check addresses the method of estimation: linear probability models. A worry might be that the event we study (obtaining a cum laude) is relatively uncommon, and so a linear probability model might not fit the data very well. For this reason the key model was re-estimated using logistic regression. Figure S10 presents a replication of the first models estimating the main effects. Here it becomes clear that the findings are very similar (and sometimes even larger) than those estimated in linear probability models. Female PhD students have about 60% lower odds of obtaining "cum laude" than male candidates, and this predicted difference increases over the models. It is important to note that the final model—the one with advisor fixed effects—is estimated over a different sample than the linear probability model. In a logit fixed effects model, only advisors can be included that have variation in the dependent variable; that have at least one PhD candidate that has a "cum laude" distinction. This reduces that sample to about 2,490 compared to the similar model with a linear probability regression. Irrespective, the findings are qualitatively highly stable: even within the same advisor, men are much more likely to obtain cum laude than women.

Figure S10: Gender gap in 'cum laude' (logistic regression estimates)



## References

- [1] LNVH, “Monitor Vrouwelijke Hoogleraren 2019,” tech. rep., LNVH, Utrecht, 2020.
- [2] J. D. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton university press, 2009.
- [3] P. D. Allison, R. A. Williams, and V. Hippel, “Better predicted probabilities from linear probability models,” *Statistical Horizons*, 2020.
- [4] S. J. Ceci and W. M. Williams, “Understanding current causes of women’s underrepresentation in science,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 8, pp. 3157–3162, 2011.