Gender-equal funding rates conceal unequal evaluations

Thijs Bol [a], Mathijs de Vaan [b], Arnout van de Rijt [c,d,*]

[a] *Department of Sociology, University of Amsterdam, Amsterdam, the Netherlands*
[b] *Management of Organizations Group, Haas School of Business, UC Berkeley, USA*
[c] *Department of Political and Social Sciences, European University Institute, Florence, Italy*
[d] *Department of Sociology, Utrecht University, Utrecht, the Netherlands*

ARTICLE INFO

ABSTRACT

Previous studies have found little or no systematic differences in the rates at which female and male scientists are awarded funding in international grant competitions. However, past investigations have only studied outcomes, not the preceding scoring and selection process. We propose that common grant review practices–such as panel deliberations, score binning, and interview assessments–allow unequal evaluations to be corrected while staying within a framework of merit-based review. We analyzed unique data from a large funding competition, the Netherlands' Organization for Scientific Research's Talent Program, including reviewer and panel evaluation scores of both funded and unfunded proposals. We replicate prior research demonstrating gender equity in funding outcomes. At the same time, we find that men received higher evaluation scores, consistent with our argument. This gender difference is counteracted by panels funding women with lower scores than men's, redistributing 64 million euro back to women that would otherwise have gone to men. Our study thus reveals that female scientists are more poorly evaluated than their male counterparts in spite of what equality in outcome statistics might suggest.

1. Introduction

Science funding organizations, using peer review for decision-making, by-and-large achieve gender-equal funding rates[1] and 21st century success chances of female and male applicants in these competitions have been found equal across a range of studies covering different nations and scientific disciplines (Albers, 2015; Beck et al., 2017; Bornmann et al., 2007; Boyle et al., 2015; Hosek, 2005; Jayasinghe et al., 2003; Van der Lee and Ellemers, 2015; Ley and Hamilton, 2008; Marsh et al., 2008; Mutz et al., 2012; Sandström and Hällsten, 2007; U.S. Government Accountability Office, 2015; Volker and Steenbeek, 2015; Waisbren et al., 2008; Ward and Donnelly, 1998).

One interpretation of this gender equality in funding is that contemporary peer review in science funding exhibits no gender bias which implies that women's grant proposals are found equally strong as men's (Albers, 2015; Ceci and Williams, 2011; Volker and Steenbeek, 2015).[2] This has led some to argue that efforts to address underrepresentation of women in science should shift focus from scrutinizing

academic evaluation toward exclusively encouraging participation (Ceci and Williams, 2011; Williams and Ceci, 2015).

We argue that gender-equal funding rates may conceal reviewer preferences for proposals written by men. Science funding organizations find themselves caught in between two potentially conflicting demands. On the one hand, they are committed to a system that promotes a meritocratic allocation of funds, thereby funding scientists whose research accomplishments and proposals are evaluated highest (Ginther and Heggeness, 2020). On the other hand, they are under pressure by governments, watchdogs, media, and academia to meet a societal expectation of equal funding chances for women and men, particularly in STEM research (European Research Council, 2019; Van der Lee and Ellemers, 2015; U.S. Government Accountability Office, 2015). If the evaluation of applications results in gender equal funding rates, the two goals are aligned, and any organizational intervention aimed at achieving gender equality would be redundant. However, if men are evaluated more favorably than women, the two goals are conflicting. Intervention is then needed to achieve gender equality but doing so may

---

* Corresponding author at: Department of Political and Social Sciences, European University Institute, Florence, Italy.
  E-mail address: arnout.vanderijt@eui.eu (A. van de Rijt).
[1] A small number of studies do find gender inequality in funding chances (Witteman et al., 2019) or gender inequality in the size of the grant (Oliveira et al., 2019).
[2] This definition of gender bias is obviously silent on the true quality of proposals, which is inherently subjective and difficult to measure. If men are more eager to participate in competitions (Niederle and Vesterlund, 2007), women's proposals may be expected to be stronger.

reduce the perception of the process being meritocratic.

In order to meet both demands, we conjecture, review panels identify opportunities in evaluation procedures that allow adjustment of review scores. Examples include subjective assessments of candidate interview performances, selective panel discussion of strong and weak points in candidate files, and the quality binning of proposals into poor, good, and excellent, allowing gender to act as tiebreaker for the middle bin. Such conceivable interventions could mask reviewer preferences for applications from men that would otherwise be revealed through unequal funding rates.

Accordingly, our research question is: "Do gender-equal funding rates truly reflect equal evaluations or are they instead achieved through panel interventions?" To this end, we analyze unique data confidentially shared with us by the Netherlands Organization for Scientific Research (NWO). The data pertain to NWO's Talent program, or Innovation Research Incentives Scheme (IRIS). This program is the primary funding source for early- and mid-career scientists in the Netherlands and has since 2002 awarded 2.6 billion Euros to individual researchers (Appendix A). The data contain reviewer and panel evaluations at all stages of the selection process, allowing us to analyze how assessments are used to select candidates for funding.

Our core analytic strategy is four-fold: (a) verifying the existence of equal funding probabilities of women and men, (b) assessing the gender of winners and non-winners near the funding threshold, where a discontinuity involving predominantly men right below and women right above the threshold is taken to signal panel intervention, (c) evaluating whether men needed better review scores in order to advance to the interview round, with an affirmative answer again signaling panel intervention, and (d) a comparison of academic output between male and female grant winners. We triangulate these findings with data from a different funding context for which less detailed data is available, namely the ERC's Starting, Consolidator and Advanced grant competitions.

2. Literature and theory

A large number of studies find that men are more positively evaluated than women in a variety of academic settings. For example, male elite scientists in biomedical research have been found to recruit fewer women than men, even though a majority of doctoral recipients in biology-related fields are female (Sheltzer and Smith, 2014). Among academics in the humanities there is a stable gender gap in earnings that is driven by the differential promotion chances of male and female scholars (Ginther and Hayes, 2003). Female job applicants have been found only half as likely to receive excellent recommendation letters as male applicants, regardless of the gender of the letter writer (Dutt et al., 2016), and when they interview, hiring committees consider husbands' but not wives' jobs an obstacle (Rivera, 2017). In online teacher evaluations male professors are more likely to receive extremely positive ratings and comments (Storage et al., 2016), especially when rating systems provide sufficient granularity at the positive extreme (Rivera and Tilcsik, 2019).

When merit is controlled, evaluations often continue to be unequal. Science faculty have been observed to rate the same job application materials more favorably when the applicant has a male name (Moss-Racusin et al., 2012) and are more likely to respond to fictional prospective students seeking to discuss research opportunities when these students were given male names (Milkman et al., 2015). Similarly, studies have found that abstracts from male authors were evaluated to be of greater scientific quality than when those same abstracts were submitted by female authors (Knobloch-Westerwick et al., 2013; Rossiter, 1993), but see Borsuk et al. (2009). Heggeness et al. (2016) compare the recipients of NIH programs to the relevant labor market, and find evidence for a leaking pipeline in research funding: women are overrepresented in junior programs (e.g., for mentoring programs or post-doctoral positions), but underrepresented in winning independent

research grants. Finally, studies reporting gender bias such as those above are themselves evaluated less favorably by male than by female scientists, but more favorably when the same research articles are altered to report no bias (Handley et al., 2015).

Despite the substantial volume of research documenting gender inequality in evaluations, the conclusion that gender bias continues to be a major factor in the evaluation of male and female scientists in modern science is not undisputed. Tregenza (2002), the editors of Nature Neuroscience, 2006, and a recent large-scale study (Squazzoni et al., 2020) find equal acceptance rates of articles submitted by male and female scientists. Borsuk et al. (2009) find no effect of gender in a controlled experiment on the evaluation of article quality by undergraduate students. In one study that examines applications for tenure-track positions, women were found somewhat more likely to be invited for an interview and to be offered a job than men (National Research Council, 2010). And some findings of gender inequality (Budden et al., 2008) are criticized on methodological grounds (Webb et al., 2008; Whittaker et al., 2008). Ceci and Williams (2011) conclude that overall the evidence is more consistent with the null hypothesis of no gender inequality in evaluation.

While evidence on gender parity in various evaluative settings is mixed, research on science funding specifically has produced broad evidence that women and men mostly enjoy equal rates of success. In meta-analyses, no overall effect of gender on funding chances is found (Bornmann et al., 2007; Marsh et al., 2009; Mutz et al., 2012). This result holds across countries and disciplines and over time (Hosek, 2005; Jayasinghe et al., 2003; Ley and Hamilton, 2008; Marsh et al., 2008; Sandström and Hällsten, 2007; Waisbren et al., 2008). Results replicate in the contexts of STEM funding in the U.S. (U.S. Government Accountability Office, 2015), the U.K. (Boyle et al., 2015), and Belgium (Beck et al., 2017).

Two recent studies did find gender differences. Witteman et al. (2019) found that funding chances of women in NIH competitions are 0.9 percentage points lower than those of men, but this difference is attributable to lower evaluations of women as investigators, not of the proposal they write. Another study identified a discrepancy between the percentage of female applicants and female winners in the Veni competition, one of the competitions that we study (Van der Lee and Ellemers, 2015). However, critics reanalyzing their data found that this difference was attributable to competitions in some fields, like social science, having both lower funding odds and higher numbers of female applicants (Albers, 2015; Volker and Steenbeek, 2015). This led them to conclude that there was no evidence of gender inequality. More generally, gender differences in funding success may result from differences in funding available between the competitions women and men partake in, even if within any competition odds are equal (Lawson et al., 2021).

Here we propose that gender-equal funding outcomes may falsely suggest equal evaluations of applications by female and male scientists. Specifically, we argue that they may not naturally emerge from the aggregation of independent peer review scores. Rather, the formal processes of grant review adopted by funding agencies may provide procedural opportunities for upwardly adjusting weaker scores given by reviewers and panelists to the proposals of women. We distinguish four concrete mechanisms that are in principle available to panels in many funding agencies. We do not know if these mechanisms are actually used but we note that they are enabled institutionally.

First, panels often do not directly follow external reviewer evaluations in their ranking and selection of candidates. Rather, reviewer scores provide input to panel discussion, which in turn produces candidate selections. Unlike external reviewers, who are asked to individually review one or a small number of grant applications, panels decide on all applications collectively and jointly. This practice of simultaneous evaluation (Bohnet et al., 2015; Kahneman and Miller, 1986; Nowlis and Simonson, 1997) provides the global information that is needed to engineer a gender-equal funding rate through post-review re-ranking.

Second, panels may use quality categories or coarse scoring that produce broad groups of "equally" ranked candidates, for which gender can be used as a tie-breaker. Sometimes panels informally or formally categorize proposals into a small number of quality categories: excellent proposals that should absolutely advance to the next evaluation round, poor proposals that should not get funded, and a middle group of intermediate proposals (e.g., Lamont 2009). By binning applications into a small number of categories, nuanced distinctions in external reviewer averages are eliminated and gender gaps may be closed (Rivera and Tilcsik, 2019). After binning, panel discussion is dedicated to the selection of candidates in the middle bin, the grey-area. If proposals in the middle category of such a coarse classification system are thought of as having been equally evaluated, then the use of gender as tie-breaker is not perceived as in tension with merit-based review.

Third, sometimes applicants are given the opportunity to counter critical points made by reviewers in a short rebuttal. This rebuttal is then evaluated by the panel. This provides an opportunity for correcting gender bias by finding women's rebuttals stronger.

Finally, some funding organizations invite finalists to present their proposal and answer questions from panelists in person. For example, applicants for European Research Council grants travel from all over Europe to Brussels for an in-person encounter with the panel. Finalists in the NWO Talent program travel to Utrecht for an interview. The subjective and often somewhat unstructured nature of this form of merit assessment again provides wiggle room for correcting gender bias.

3. The setting: NWO's innovation research incentives scheme

In this study we analyze the Innovation Research Incentives Scheme (IRIS) funding competition from the Netherlands Organization of Scientific Research (NWO). IRIS is the primary individual funding source for early- and mid-career scientists in the Netherlands and consists of three competitions: "Veni" (recent Ph.D.s), "Vidi" (up to 8 years after PhD), and "Vici" (up to 15 years after Ph.D.). Applications are evaluated by eight domain-specific panels (e.g., medicine, social sciences, physics), who rely on reports of external peer reviewers before making a final ranking. The IRIS-scheme is the main instrument through which NWO allocates public funding to talented scientists in the Netherlands. Moreover, it is the only funding vehicle in the Netherlands that provides large grants on an individual basis. An overview of the main characteristics of the IRIS-scheme is provided in Appendix A (Table S1).

3.1. Evaluation in the IRIS-scheme

The first stage of a Veni, Vidi, or Vici-grant application process requires applicants to submit a proposal. For the Vici-grant, applicants only submit a pre-proposal. Evaluation of these (pre-) proposals takes place within eight research divisions (Appendix, Table S2). NWO-divisions are not of equal size, and the numbers of grant applications submitted to these divisions varies greatly.

The gender composition and total success rate for applicants differs considerably across the NWO-divisions as well. Overall, there are more female applicants in divisions that have a lower overall success rate (Albers, 2015). Each application is evaluated by about 10 panel members. Panels consist of Dutch scientists and are appointed by the scientific board of each NWO division. In the period under study (2005–2016),[3] the dimensions on which panels score applicants have changed. In the period 2005–2008 only two dimensions were taken into account: quality of the researcher (50%), and quality of the proposal (50%). After 2008 a third dimension (knowledge utilization) was introduced, and the final priority score was based on the quality of the researcher (40%), the quality of the proposal (40%), and the potential for knowledge utilization (20%). Panelists can assign each of these three

factors any score ranging from 1 (excellent) to 9 (unfundable). The aggregate of these factors, averaged over all panelists, constitutes the priority score of the applicant.

For all grant competitions that we study, the panel makes selection decisions at three moments: (a) pre-selection, (b) after external reviews, and (c) after the interviews. The timeline of the selection is schematically described in Fig. 1. The full process takes about 8 months and is similar across the three grant programs, although there are some differences.

For the Veni- and Vidi-competitions full proposals are submitted immediately. The first evaluation moment is during the pre-selection before applications are sent out to external reviewers. Pre-selection only takes place when the number of proposals that is submitted is over four times the number of grants that can be awarded. In the pre-selection stage, divisional panel members read the full proposals and rank them by scoring the quality of the proposal, the quality of the researcher and the quality of the knowledge utilization. The panelists each evaluate 15–25 proposals and each proposal is evaluated by multiple panelists. Proposals are then ranked on the basis of the average across panelists' scores. The maximum number of applications that is sent out to external reviewers equals four times the number of available grants.

In the second stage there are two evaluation moments. First the remaining proposals are evaluated by external reviewers (2 to 3 for each Veni-proposal and 3 to 4 for each Vidi-proposal). External reviewers are asked to assign each proposal with a letter: *A+* (excellent, 1), *A* (high quality, 3), *B* (good quality, 5), UF (good, but unsuccessful in its current form, 7), and *U* (unfundable, 9). External reviewers are selected by staff members of NWO—not by the panel—and remain anonymous to both applicants and panelists. NWO approaches external reviewers based on the abstract, key words, literature list, and library tools.

Once the reviews have been completed and NWO has provided applicants with those reviews, applicants have one week to write a two-page rebuttal in which they can reflect on issues raised by the external reviewers. The second evaluation in this stage is done by the panel members: they read the applications, the external reviews, and the rebuttals, and collectively scores all candidates during a dedicated meeting. By now, candidates have already been evaluated up to 3 times: first by panelists in the pre-selection, then by external reviewers, and finally again by panelists. If there are *N* available funding slots within a given NWO-division, the division typically invites up to 2xN applicants for the final stage in the application process: the interview.

The final moment of evaluation takes place after the final selection stage: the interviews. After the interviews with the candidates, panels make a final ranking and assign each candidate a final priority score ranging from 9 (lowest ranking) to 1 (highest ranking). All individual panelists give a score, and the final priority score for an application is the panel average. Panelists then discuss the final rankings. There is a predetermined number of grants that can be awarded, and panelists are aware of that number. Once the rankings of the NWO-divisions are finalized, the NWO-board compiles a list of winners.

Applicants to the Vici-grant competitions initially only submit a pre-proposal. These pre- proposals are substantially shorter than proposals in the Veni and Vidi competitions. The pre-proposals are evaluated by panels formed in the three domains (see Appendix, Table S2), and applicants are provided with a recommendation about whether or not to submit a full proposal (yes/no). Irrespective of this advice, everyone is allowed to submit a full proposal for the Vici. The remainder of the evaluation process in the Vici is similar to the Veni and Vidi competitions.

---

[3] See Sample for an explanation of the analytic sample of the study.

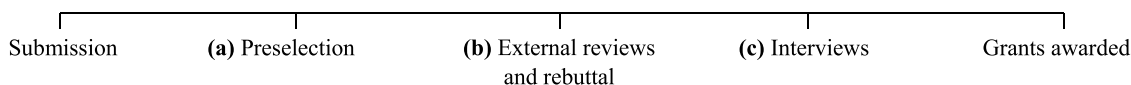| Submission	**(a)** Preselection	**(b)** External reviews and rebuttal	**(c)** Interviews	Grants awarded

Fig. 1. The selection process.

3.2. Gender policy

NWO has publicly expressed to strive for equal funding percentages of female and male applicants in each of its competitions.[4] However, panels have historically lacked and still lack the formal mechanisms for ensuring this. Panelists are reminded of the importance of gender-neutral evaluation and of taking effective research time and career breaks into account. There are no quota, however. Only in case of equal average evaluation scores do women have precedence.

4. Data

Our main dataset combines all application records kept by NWO. The data are unique in that they provide evaluation scores on both winners and non-winners at all stages of the selection process. This allows us to examine the dynamic process through which equality is achieved, not just the outcome, as in the previous studies mentioned. Moreover, because the data are identified, we can match additional academic success variables to compare the profiles of female and male grant winners. The final dataset consists of 12,555 unique applicants and 20,360 applications. Our data contain demographic information on the applicants (e.g., gender) and application-specific information (e.g., grant scheme, year submitted, final ranking, external reviewer scores, grant awarded). We merged publication and citation data from the Scopus database to grant winners in our data in order to investigate whether female and male winners perform differently with respect to academic journal publishing. An overview of the variables that we will analyze will be discussed below.

Our analytic sample contains all applicants for which we have information on their panel priority scores as well as their external reviewer scores. In the NWO archives this information is recorded for most years. Before 2005 our data do not contain the scores from external reviewers and this means that our analyses do not refer to this period. After 2005, we have information for almost all rounds and divisions. An overview of the missing years and divisions can be found in Appendix A (Table S3).

4.1. Data on applicants and evaluation

The first variable that is crucial for our data analysis is an applicant's self-reported gender identity. The percentage female applicants rises from about 30% in early years to about 40% in later years. Fewer than 10 applicants identified as neither male nor female. Because of the risk their small number poses to identification, we exclude their cases from analyses.

A second set of variables that we use in our analyses are obtained from the applications. We know for each application to which grant scheme it was submitted (Veni, Vidi, or Vici), in which year (2005–2016), and to which NWO-division.

For each application we have detailed information on evaluations at different stages of the selection process, including an indicator of whether a grant was awarded (1) or not (0), the average final priority score that panelists assigned to the proposal, ranging from 1 ("excellent") to 9 ("unfundable"), and its final ranking. For most applications in the eraly years of our data we do not know how panelist scores changed across different stages—we only have access to the final score. For more recent years, however, we have time-varying information on the priority scores that panelists assigned to the same proposals. This means that for a subset of applications (Veni 2014–2016), we can analyze how panelists changed their scores for the same proposals over the course of the grant application process.

Finally, for all proposals that made it through the preselection stage we have information on the scores that external reviewers assigned to the application. External reviewers are asked to provide a written assessment of the application and then rank it with an *A+*, *A, B, UF*, or *U*. Following NWO guidelines, we have recoded the external reviewer assessments in scores ranging from 1 (*A+*), 3 (*A*), 5 (*B*), 7 (*UF*), to 9 (*U*). Combined grades were also given by external reviewers and were assigned the midpoints: *A+/A* becomes 2 and *A/B* becomes 4.

4.2. Publication and citation data

To assess whether there are measurable quality differences between male and female applicants, we use auxiliary publication data that we extract from Scopus. Specifically, we match Scopus information to all successful applicants between 2005 to 2016 with two objectives in mind. First, it allows us to evaluate whether there are academic profile differences between female and male winners prior to winning. In the absence of an objective measure of application quality, we use pre-competition publication records as an imperfect proxy. The second objective is to be able to assess whether female winners went on to have more or less output and recognition than male winners.

We wrote a program that interacts with the Scopus API and returns a list of potential Scopus IDs for a combination of last names and first names from our data. When the first name is missing, we use initials. For about 51% of the cases only a single match is found. For the remainder of the cases, multiple Scopus IDs are returned. Often this is caused by single authors having multiple IDs in Scopus (e.g., Tom de Bruin and Tom G. de Bruin). If this was the case, different Scopus profiles for the same authors were combined manually. When there were multiple scholars with the same name (i.e., David Smith), we used information on gender, research topic (derived from the name applications), and home institution to manually establish the correct match. For about 169 applicants, the program did not find any match. This was often caused by a different spelling of the last name in the NWO-files compared to the Scopus database. For these cases, we added the Scopus ID manually. In total we were able to match a Scopus ID for 2,374 applicants that won a grant and had no missing information on their evaluation scores. We were unable to find reliable information on only 13 applicants (0.9%).[5] They are removed from the analyses that use Scopus data.

We use three measurements from the Scopus database: (a) number of publications, (b) average number of citations per publication, and (c) *H*-index. We obtain these three measures for grant winners before they obtained the grant and for the four years following their grant win. Building on these data we are able to evaluate if male and female applicants' academic output was similar before obtaining the grant and after winning the grant. These measures of academic output have clear limitations. We do, for example, not know whether applicants wrote academic books that received citations. To mitigate this concern, we compare applicants from the same academic domain: we do not compare female sociologists to male physicists precisely because publication cultures are so different. Still, gender differences in publishing cultures within academic disciplines (i.e., female sociologists publishing more

books than male sociologists) are not accounted for. An overview of all the main variables that we use in our analyses can be found in Table 1. Descriptive statistics can be found in Appendix A (Table S4).

5. Results

5.1. Gender differences in grant evaluations

Fig. 2 compares the percentages of women among grant applicants and winners in the Veni, Vidi, and Vici competitions. The results replicate the general result of gender equity in funding chances that is found in the literature for peer-review-based grant competitions worldwide. In the Veni competition women overall exhibit slightly lower funding chances (Van der Lee and Ellemers, 2015) (Pearson χ^2 test, $p < 0.05$). Consistent with earlier analyses (Albers, 2015; Volker and Steenbeek, 2015) we find that when breaking down the Veni results by field there is no significant gender difference in the probability of receiving a Veni grant (logistic regression, $p \geq 0.05$ in Appendix C, Table S6). For the Vidi and Vici competitions (Fig. 2) we find no statistical gender differences in the percentages of grant winners (Pearson χ^2 tests, $p \geq 0.05$), and the lack of difference remains when controlling for the scientific field of an application (logistic regression, $p \geq 0.05$ in Appendix C, Table S6).

We next evaluate whether this proportional representation of women among grant winners accurately reflects equal evaluations of proposals by male and female applicants. Fig. 3 depicts the male-female gap in external reviewer scores of all applicants, the gap for just those applicants who were invited for an interview by the panel, and the gap in the final panel scores of interviewees on the basis of which funding decisions were made. Strikingly, in the Veni, Vidi and Vici competitions male applicants consistently received better external reviewer scores (2-sample t-test; $p < 0.01$ in each case).

A potential mechanism that might explain why female scholars receive lower external reviewer scores is that male reviewers are more critical of female applicants than are female reviewers while no such relationship exists for male applicants. To examine whether reviewer gender might moderate the effect of applicant gender on reviewer score, we estimated regression models in which we interact the gender composition of the reviewers with the gender of the applicant (details in Appendix D). In these additional analyses we do not find any evidence that the gender composition of the reviewers affects the gender gap in evaluations. This result concords with most prior research on the role of evaluator gender in allocative settings suggesting that reviewer gender

Table 1
Overview of main variables.

| Variable	Description
NWO data	
Gender	Gender of the applicant.
NWO-divisions	The division of the applicant (see Table S2).
Grant round	The round of the grant application (2005–2016).
Veni winner	Binary variable indicating whether the applicant won the Veni grant (0 = no, 1 = yes).
Vidi winner	Binary variable indicating whether the applicant won the Vidi grant (0 = no, 1 = yes).
Vici winner	Binary variable indicating whether the applicant won the Vici grant (0 = no, 1 = yes).
Panel score	The final priority score (1–9) that a panel assigns to each application in the grant competition.
Rank	The final rank within the grant, round, and NWO division of applicants. This rank is based on the panel score.
External reviewer score	The within-candidate mean of the external reviewer scores. Reviewers can score candidates from $A+$ (1) to U (9).
Scopus data	
Publications	Count variable that captures the number of publication in a given year.
Citations	Average number of citations per publication.
H-index	An index of h implies that a scholar has published h papers each of which has been cited in other papers at least h times.
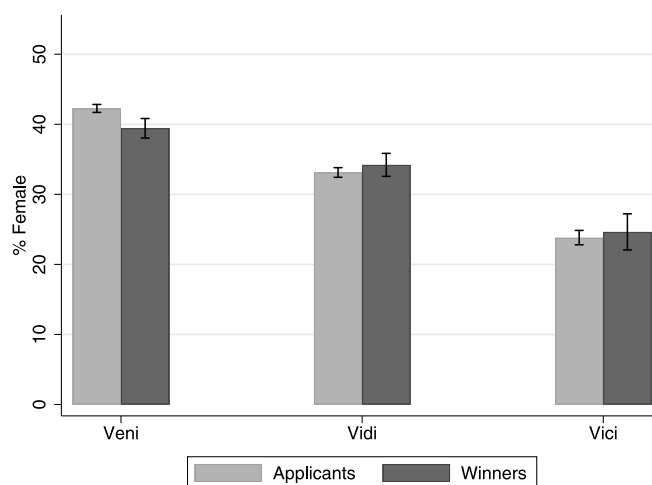
Fig. 2. Share of women among applicants and winners
Note. Sample contains all grant applicants to the Veni, Vidi, and Vici in the years 2005–2016. Percentages of female applicants and winners in the Veni, Vidi and Vici science funding competitions. Whiskers depict standard errors of the estimates.

is unlikely to undergird any gender gap in evaluation (Bagues et al., 2017; Goldin and Rouse, 2000; Ridgeway and Correll, 2004).

The gender gap in external review scores revealed in Fig. 3 persisted and remained significant ($p < 0.01$ in each case) among those invited for an interview, indicating that the female applicants that panels invited for an interview on average had lower review scores than the male applicants they invited. This suggests that the panel discussions of external reviews and applicant rebuttals produced a commensurate upward adjustment in panel evaluation scores for women resulting in more women invited to the interview.

While there is a clear gender gap in the scores of the external reviewers, Fig. 3 also shows that the final scores given by panels did not differ between female and male interviewees ($p \geq 0.05$ in each competition). This suggests that panels applied commensurate counterforce in their selection of interviewees and final evaluations, attempting to rebalance the distribution of gender. In the Vidi (0.18sd) and Vici (0.34sd) competitions reviewer scores were even more strongly in favor of male applications. These findings are robust when controlling for differences in external reviewer scores across panels and divisions (Appendix C, Table S7). Moreover, there are no systematic differences across divisions with more or less female applicants, suggesting that the gender gap may not be a function of how gender stereotyped a field is.

As discussed in our introduction, studies have found that evaluators are less subject to gender bias when they must rate multiple candidates rather than just one in isolation (Bohnet et al., 2015; Chang et al., 2019; Kahneman and Miller, 1986; Nowlis and Simonson, 1997). This raises the possibility that the discrepancy between panel and reviewer evaluations in Fig. 3 is due to panelists comparing applicants while reviewers are asked to rate only one. We evaluate whether it is plausible that the simultaneous evaluation of a larger number of candidates was an important driver for the convergence of scores of male and female applicants. We do so by analyzing auxiliary data from a selection of Veni competitions where panel scores before the external reviews were available.

Fig. 4 presents the marginal predicted gender gap from a regression model where we controlled for divisional effects. In the preselection (see Fig. 1), panelists initially evaluated female applicants to the Veni grant lower than male applicants: a gap of about .18sd. At the end of the evaluation cycle, after the interviews in round 3, the gender gap in the same group of candidates disappears. This difference in the gender gap between round 1 and round 3 is 0.17sd and statistically significant ($p < 0.05$). This speaks against simultaneous evaluation of multiple
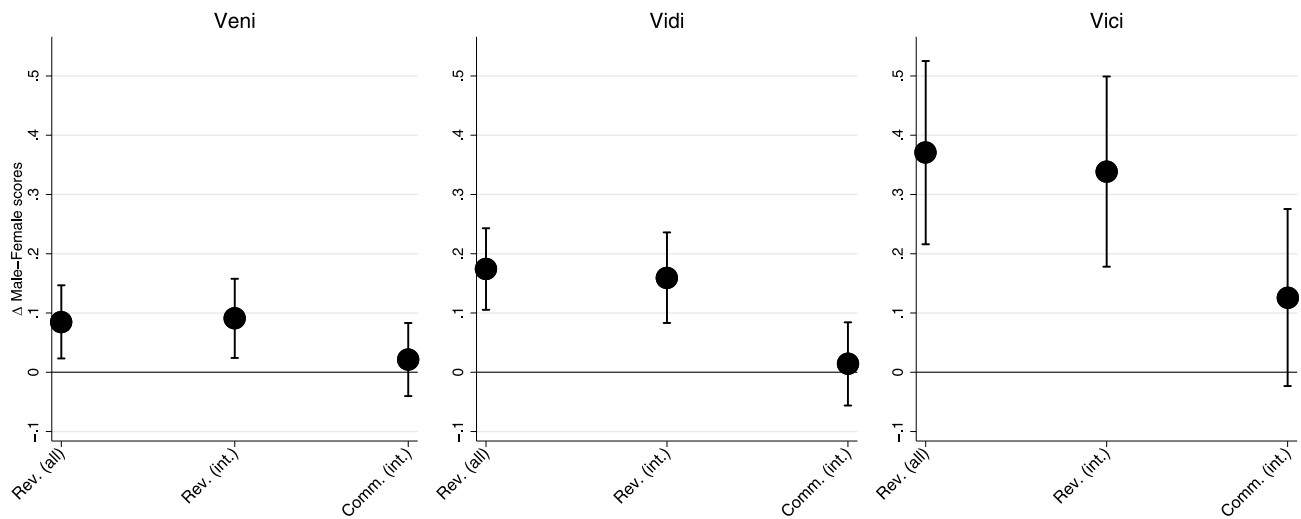
Fig. 3. Difference between female and male applicant scores of reviewers and panels
Note. Calculations based on a sample that contains all applicants that progressed to the external reviewer round and did not have missing data on external reviewer scores (Appendix A). The sample is smaller for the second and third marker, as these only contains candidates that progressed to the interview round. Markers represent the male advantage in the Veni, Vidi, and Vici competitions, calculated by normalizing scores across the three grant programs and subtracting the average female score from the average male score. Shown are male advantages in external reviewer scores across all applications ("all"), across just those applications that proceeded to the interview stage ("int.") and in final scores given by panels. Whiskers depict 95% confidence intervals.
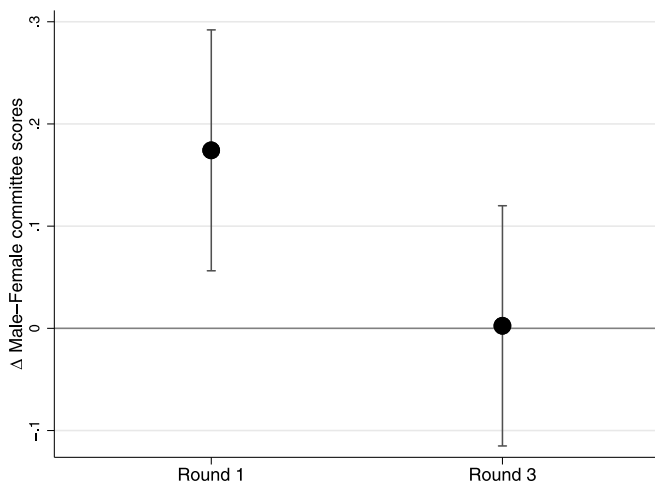

Fig. 4. Difference between female and male applicant scores of panels in different rounds
Note. Markers represent the male advantage in the Veni competitions from 2014 to 2016, obtained from models where division differences are taken into account. Shown are male advantages in z-scores before the external reviewer reports (Round 1) and after the final interview (Round 3). Whiskers depict 95% confidence intervals.

candidates being responsible for the convergence in scores between male and female applicants, because the evaluation procedure was simultaneous in both round 1 and round 3.

Perhaps the clearest evidence for deliberate intervention comes from an analysis of final panel ranks. Men who advanced to the final stage on average had higher external reviewer scores (Fig. 3), which makes an intervention in the final ranking of candidates necessary if gender parity in funding is to be achieved. The most effective procedure for downward adjustment of scores (Fig. 4) would be to swap male and female candidates around the threshold after their interviews. Fig. 5 shows the percentage of female applicants with a given rank away from the funding threshold for each competition stage. Negative ranks (-1, -2, -3) indicate applicants just below the funding threshold, positive ranks (1, 2, 3) those just above it. The percentage of women tends to dip right

below the funding threshold and peak right above it, indicating that in this grey area of evaluation applicant gender was used as a scoring criterion by panels. This non-monotonic relationship between gender and final rank is visible in each of the three competition stages. In the Veni and Vidi competition the jump in the percentage of women from rank -1 to rank $+1$ is statistically significant (Chi square test, $p < 0.05$). The Vici competition shows a comparable increase from rank -1 to rank $+1$ but no significant difference is found ($p \geq 0.05$). Together, the results provide evidence that corrective action by panels counteracted the effects of better male external review scores, thereby producing the roughly proportional representation of women among grantees seen in Fig. 1 and found in earlier studies (Albers, 2015; Van der Lee and Ellemers, 2015; Volker and Steenbeek, 2015).

5.2. Implications of gender differences in grant evaluations

To quantify the impact of panelists' deviations from reviewer scores, Table 2 shows the number of women who would not have been funded had external reviewer preferences for male applications not been counterbalanced. This number is computed by ranking applicants based on their review scores, computing the number of women among the N highest-ranked applicants and subtracting this from the number of women among the N actual winners. Because applications around the Nth rank can have tied review scores, we base our estimates on 1000 simulations in which we randomly ranked applications with equal review scores. In addition to the number of women winning in the counterfactual scenario, Table 2 also shows the euro amount that women would not have received had reviewer skew toward male applicants not been counterbalanced. Table 2 depicts that, had panels not intervened, 16 (3%) of the 485 female Veni grantees would not have been funded and 4 million euros would have gone to men instead of women. Impact was progressively greater in later-stage competitions, with 33 (12%) of all 284 female Vidi grantees being due to panel intervention, and 26.7 million euros extra going to research done by women. In the Vici competition, 22 (32%) of all 69 female laureates received their grant because panels were more positive about applications by women than reviewers were, redistributing 33 million euros from male to female applicants.
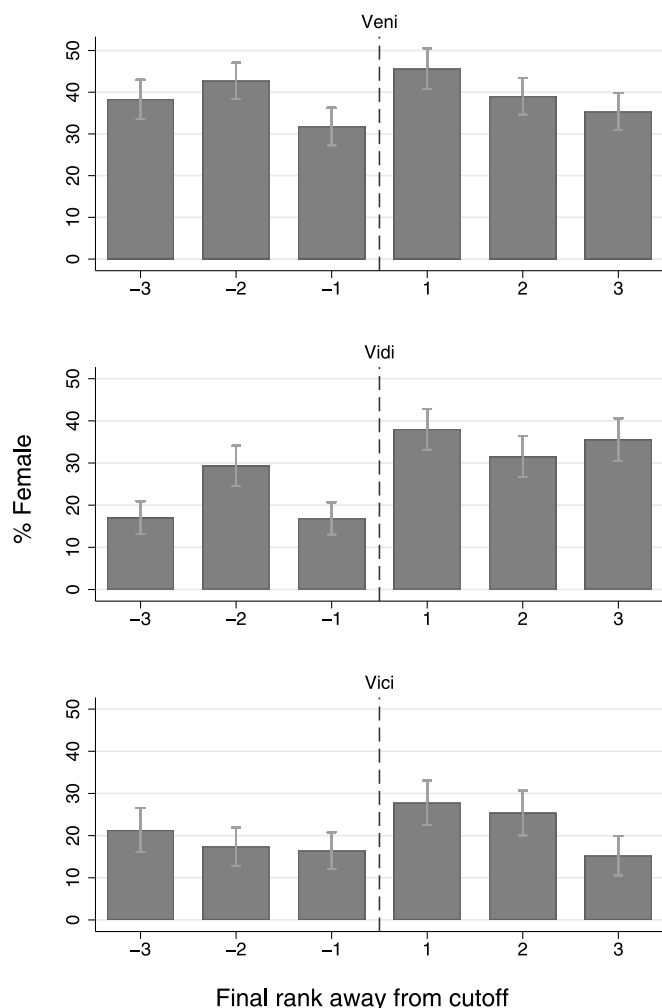
Fig. 5. Share of women near the funding threshold
Note. The bars in the Figure are based on the applicants just around the
threshold for all competitions in the years 2005–2016 for which we have ob-
tained data. Shown is the average percentage of women across competitions
with a given rank away from the funding threshold. The dashed line represents
the cutoff. Whiskers depict standard errors.

Table 2
Impact of panel intervention on female competition success.

	Veni	Vidi	Vici
# Female grantees	485	284	69
# Female grantees otherwise not funded	16	33	22
Euros otherwise not granted to women	€ 3890,000	€ 26,710,000	€ 32,940,000

5.3. Gender differences in academic output

In a final set of analyses we ask whether female winners had equally
strong profiles as male winners. If the lower evaluation scores of female
applicants found in Figs. 3 and 4 are based on weaker candidate profiles,
then the panel interventions should have produced a pool of winners in
which women on average have fewer publications and citations than
men. To test this, we obtained publication and citation counts from the
Scopus database for the 98% of grant winners whom we could unam-
biguously match on name, field and institutional affiliation.

In Fig. 6, we evaluate whether male and female Veni, Vidi, and Vici
winners with similar final evaluation scores had comparable academic
performance in the years pre and the 4 years post grant competition. To

ensure that we are comparing men and women who had received similar
final evaluation scores, we only include a winner i at rank x if there is an
opposite gender winner j at rank $x-1$, x, or $x+1$. Doing so prevents us
from comparing male and female winners whose differences in aca-
demic performance are already reflected in their rankings.[6] In Fig. 6, we
compare the numbers of citations, publications and H-indices of female
and male grantees before and after the Veni, Vidi, and Vici competitions.
Positive scores denote a male advantage.

The graphs show that among winners in the Veni competition,
women were better cited prior to receiving the award (paired t-test; $p <$
0.05) but had fewer publications than men ($p < 0.05$ in the 4 post-grant
years), while there was no difference in H-index ($p \geq 0.05$). These dif-
ferences are small to moderate: In the social sciences, for example, these
standardized effects translate into women having on average slightly
less than 1 citation per article more than men (a difference of about 10%
relative to the mean) and men having on average slightly less than 1
publication more than women (a difference of about 9% relative to the
mean). The same pattern is visible during the years after winning the
Veni grant. In the Vidi competition, women and men have comparable
citation rates, while men publish slightly more, both before and after the
competition. In the Vici competition, men seem to outperform women
on all three indicators, although very few of them reach conventional
levels of statistical significance: only for number of publication prior to
winning the award we find a significant difference ($p < 0.05$). The low
number of observations in the Vici competition limits power. When
Veni, Vidi, and Vici are pooled (Appendix E), men are found to have
significantly more publications and a significantly higher H index before
and after the grant, while there is no difference in citations between
male and female winners either before or after the grant. The differences
remain small, however: the higher number of publications for men
equals only 0.18 standard deviations of the distribution of publication
totals. At the same time we find no gender differences in the change in
either of the performance indicators when comparing the pre-grant with
the post-grant period (Appendix E).

Altogether, the evidence suggests that panels awarded grants to
women and men with reasonably comparable profiles, with perhaps a
publication penalty for men.

The evidence presented in Fig. 6 should be interpreted with an eye to
a number of considerations. First, the analysis leaves out other impor-
tant dimensions of accomplishment. For example, the analysis is silent
on the quality of women's and men's proposals. Second, the outcome
data included in the analysis may not accurately reflect the quality of
work of women in case of gender bias that preceded the grant applica-
tion process. While much of the recent research on gender bias in the
publication process suggests that conditional on submission, male and
female scholars are equally likely to publish their research (Forscher
et al., 2019; Squazzoni et al., 2020), other – especially older – work
suggests that throughout the publication process contributions of female
scholars have been systematically undervalued (Tregenza, 2002; Wold
and Wennerås, 2010). One explanation for this contrast in findings is
that awareness of gender bias and changes in policy and behavior have
allowed the publication process to become more equal over time
(Roper, 2019). But regardless of such changes, unequal career oppor-
tunities upstream of the grant application processes may have made it
harder for women to publish and be recognized resulting in publication
and citation records favoring men.

5.4. Generalizability

An important limitation of our main analysis is that it is restricted to
a single funding organization, raising questions of generalizability.

---

[6] This strategy is effective in creating samples of winners with similar eval-
uation scores as t-tests of differences in either ranking or scores between men
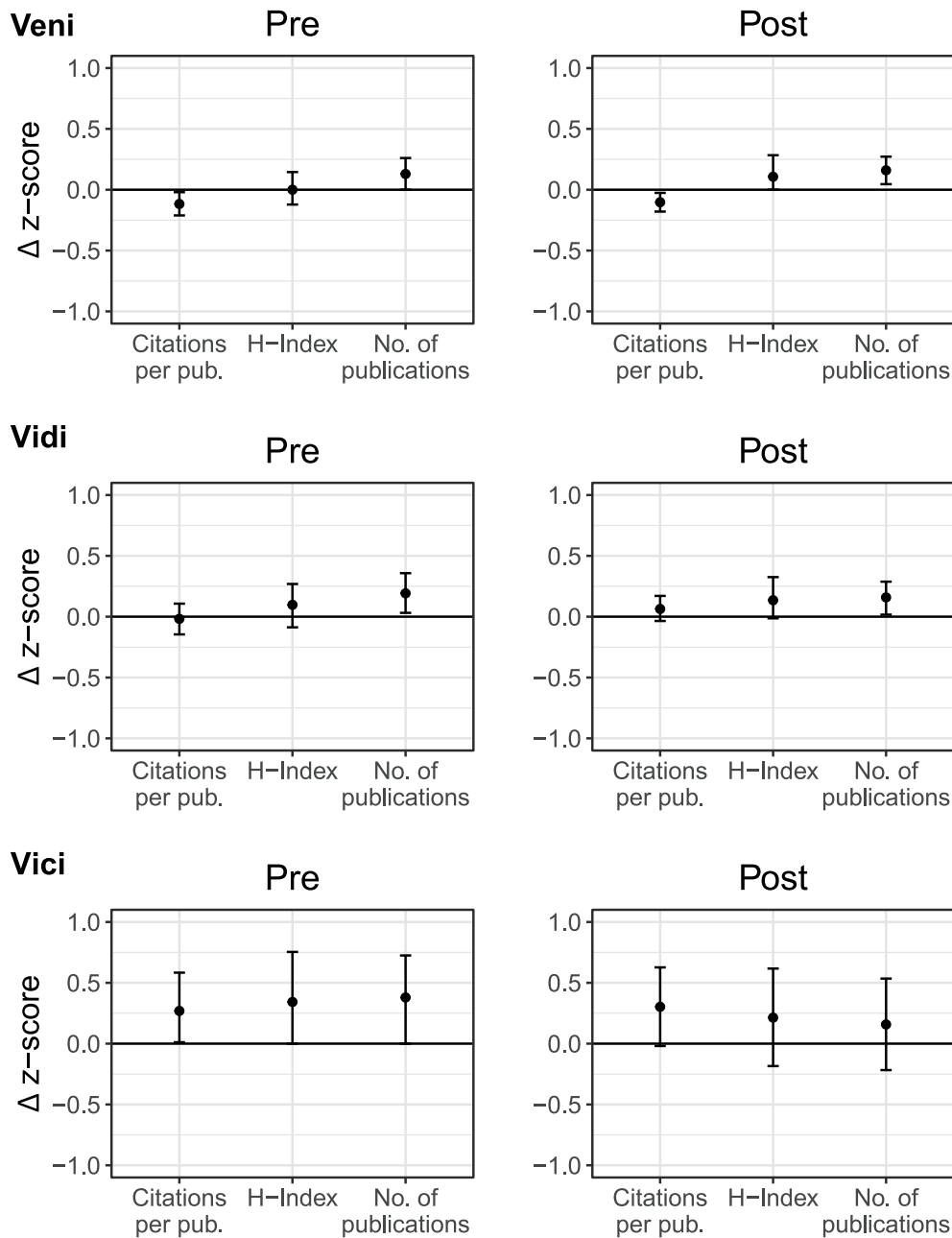and women reveal no significant differences.

Fig. 6. Gender differences among winners in citations, H-index, and publications
Note. The analytical sample Citations, *H*-index, and number of publications are standardized by field and year. Pre and post denote the periods during the 5 years preceding and following the Veni, Vidi, and Vici competitions. Positive values denote a male advantage. Whiskers depict 95% confidence intervals.

Other competitions are under similar scrutiny by the general public and the scientific community (Edlund, 2018; European Research Council, 2019; National Research Council, 2010; U.S. Government Accountability Office, 2015) and the widespread use of panel discussions and broad categories of funding priority leave program directors and panel chairs substantial room to engage in compensatory practices such as the ones observed in the Dutch case here. We had an opportunity to explore whether similar results could be found using more limited data from a different funding organization. The European Research Council's Starting (€1,500,000), Consolidator (€2,000,000), and Advanced (€2,500,000) grants form the largest European competition for individual researcher funding with an annual budget of a little under 2 billion euros. We obtained data on the applicant gender and evaluation scores of all 22,279 applications submitted to these three competitions during the years 2014–2016. In the ERC evaluation process, as in the NWO

procedures, panels first make a preselection of proposals that proceed to the subsequent round. In this second round, external reviewers and panelists determine a final grade, which is either an *A* or *B*. Only a portion of the proposals that are given an *A* are then funded (Edlund, 2018). Our ERC data are clearly more restricted than the NWO data, with less granularity on scores and anonymized applicants, limiting possibilities for analysis. Still, they allow us to probe whether also here men are favored early in the selection process, while this gender difference is corrected in later rounds. Specifically, we have three binary success variables: (i) whether an application advanced to round 2, (ii) whether an application received an *A* grade necessary for funding, and (iii) whether the application was funded. We estimate conditional logistic regression models for each outcome variable, with as strata panels (e.g. PE09, meaning physics and engineering panel nr. 9) nested within calls (Starting / Consolidator / Advanced). As such we made sure that

only proposals evaluated by the same panel were compared.

Fig. 7 shows odds ratios of women's success relative to that of men at the various stages in the evaluation process. The overall odds of funding do not vary by gender (left of the dashed vertical line). However, just as they did for NWO, these equal funding odds conceal gender differences in success at different stages (right of the dashed line). In Round 1, applications from women fare worse than those of men, as they also did in the preselection stage of the NWO competition. This leads fewer female proposals to be sent out for external review in Round 2. In Round 2, the preselected proposals of women and men achieve equal grades, suggesting that external reviewers agree that the lower preselection odds for women were meritorious. Finally, of those applications receiving an A, proposals by women are 40% more likely funded by panels. Without the last round correction, an additional 8.3% of the total €5400,000,000 distributed in the 2014–2016 calls, or €450,000,000, would have gone to men instead of women. These results bear striking similarity to the NWO results, with women in early rounds being evaluated more poorly–both by panelists and reviewers–only to receive favorable treatment at the final decision stage, so that gender parity is nonetheless achieved.

6. Discussion

We asked if women and men's equal chances of winning a grant reflect similarly equal peer evaluation. Our analysis of unique data on proposal evaluations suggests they do not. Peer review did not seem to naturally produce equal funding chances for women and men. The funding organization we studied does produce comparable funding rates for female and male applications. However, behind the veil of equal funding outcomes, evaluators show a preference for male applications over female applications. The gender gap in scores is substantial: up to 40% of a standard deviation in a grant competition for mid-career researchers. The reason this hidden difference does not translate into worse funding chances for women appears to be that panels rectify it in their funding decisions, thereby preventing gender imbalance among awardees.

The rectification occurs in two steps: First, men on average need higher external reviewer scores in order to advance to the interview
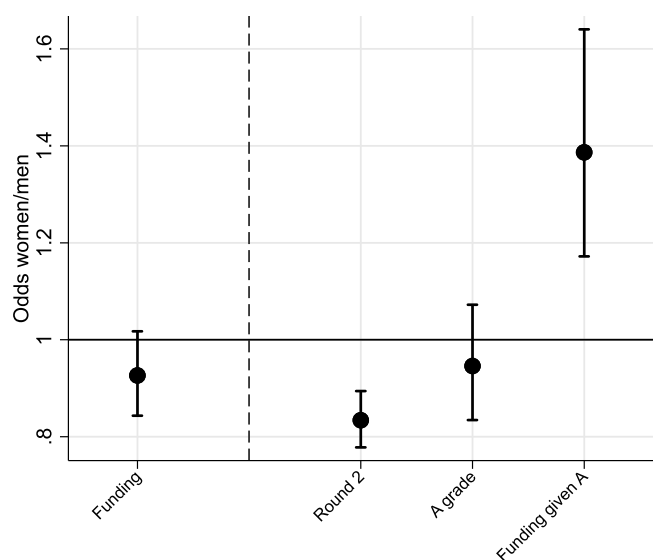

Fig. 7. Gender differences in ERC application evaluation
Note. Odds-ratios of women vis-à-vis men receiving ERC funding, proceeding to round 2, receiving an *A* grade in round 2, and receiving funding given an *A* grade in round 2 (*N* = 22,279). Estimates of odds-ratios with their confidence intervals from logistic regression models with fixed effects for call (starting, consolidator, advanced) and domain panel (e.g. "climatology and climate change").

round. This is achieved in the rebuttal stage of the evaluation process where we speculate that the debate between applicant and reviewer on merit provides legitimacy for the panel's departure from reviewer opinion. Second, at the very end of the evaluation process panels swap women and men with similar evaluation scores near the funding threshold, thus achieving equal funding chances. Legitimacy for this intervention is likely found in panel discussion about candidate performance in the interview and through the use of coarse quality categories that allow a male candidate with somewhat better scores to share a quality category with a female candidate who can then be favored on the basis of a tie breaker. While we observe the results of the decisions made by evaluators, we do not directly observe how they arrive at these decisions. A source of data that we do not have access to but could potentially reveal some of the details of these decision-making processes are the evaluation reports generated by reviewers. Future work may use such data to further illuminate on how funding organizations navigate issues of meritocracy and equality.

Our results also demonstrate that even though external reviewers were exposed to nudges aimed at addressing gender bias, these nudges did not produce gender-equal evaluations. Prior work has found that other policies aimed at reducing implicit bias (e.g. diversity training) have limited effects on behavioral change (Bezrukova et al., 2016; Chang et al., 2019). While we do not know how unequal evaluations would have been without the nudges, the results suggest that organizations aiming to reduce the effects of implicit biases should be realistic about what can be achieved through mere training and campaigns.

Reviewers were overwhelmingly male. Nonetheless, we found that female reviewers did not exhibit a stronger average preference for applications from female applicants than male reviewers did. This is consistent with findings from prior work on evaluator gender in arts and science (Bagues et al., 2017; Goldin and Rouse, 2000). It also suggests that increasing female participation in grant peer review is not likely to be an effective policy for mitigating gender differences in funding.

Our results indicate that female and male winners in the competitions we studied did not differ greatly in terms of their publication and citation profiles. This suggests that organizations can reasonably succeed at accommodating gender parity in funding within a merit-based review process. However, this gender parity does not emerge naturally but instead is produced through panels' exploitation of opportunities for corrective intervention in the evaluation process. Our study thus reveals that female scientists are more poorly evaluated than their male counterparts in spite of what equality in outcome statistics might suggest.

CRediT authorship contribution statement

Thijs Bol: Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Investigation, Data curation, Visualization, Formal analysis. **Mathijs de Vaan:** Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Investigation, Data curation, Visualization, Formal analysis. **Arnout van de Rijt:** Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics

This project was approved by the Institutional Review Board of the Amsterdam Institute for Social Sciecne Research at the University of Amsterdam. The project was conducted in cooperation and in compliance with data access restrictions imposed by the Netherlands Organization for Scientific Research (NWO) and the European Research

Council (ERC). The ERC data were anonymous. The NWO funding agency data were anonymized after name-based matching with publication and citation data from SCOPUS.

Acknowledgements

References

Albers, C.J., 2015. Dutch research funding, gender bias, and Simpson's paradox. Proc. Natl. Acad. Sci. 112 (50), E6828–E6829. https://doi.org/10.1073/pnas.1518936112.

Bagues, M., Sylos-Labini, M., Zinovyeva, N., 2017. Does the Gender composition of scientific committees matter? Am. Econ. Rev. 107 (4), 1207–1238.

Beck, R., Halloin, V., 2017. Gender and Research Funding Success: case of the F.R.S.-F. Belgian. Res. Eval. 26 (2), 115–123. https://doi.org/10.1093/reseval/rvx008.

Bezrukova, K., Spell, C.S., Perry, J.L., Jehn, K.A., 2016. A meta-analytical integration of over 40 years of research on diversity training evaluation. Psychol. Bull. 142 (11), 1227.

Bohnet, I., van Geen, A., Bazerman, M., 2015. When performance trumps gender bias: joint vs. separate evaluation. Manag. Sci. 62 (5), 1225–1234. https://doi.org/10.1287/mnsc.2015.2186.

Bornmann, L., Mutz, R., Daniel, H.D., 2007. Gender differences in grant peer review: a meta-analysis. J. Informetr. 1 (3), 226–238.

Borsuk, R.M., Aarssen, L.W., Budden, A.E., Koricheva, J., Leimu, R., Tregenza, T., Lortie, C.J., 2009. To name or not to name: the effect of changing author gender on peer review. Bioscience 59 (11), 985–989.

Boyle, P.J., Smith, L.K., Cooper, N.J., Williams, K.S., O'Connor, H., 2015. Gender balance: women are funded more fairly in social science. Nat. News 525 (7568), 181. https://doi.org/10.1038/525181a.

Budden, A.E., Tregenza, T., Aarssen, L.W., Koricheva, J., Leimu, R., Lortie, C.J., 2008. Double-blind review favours increased representation of female authors. Trends Ecol. Evol. 23 (1), 4–6 (Amst.).

Ceci, S.J., Williams, W.M., 2011. Understanding current causes of women's underrepresentation in science. Proc. Natl. Acad. Sci. 108 (8), 3157–3162.

Chang, E.H., Milkman, K.L., Chugh, D., Akinola, M., 2019. Diversity thresholds: how social norms, visibility, and scrutiny relate to group composition. Acad. Manag. J. 62 (1), 144–171.

Dutt, K., Pfaff, D.L., Bernstein, A.F., Dillard, J.S., Block, C.J., 2016. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. Nat. Geosci. 9 (11), 805.

P. Edlund 2018. Constructing an arbiter of status: the european research council's emergence in the field of science. PhD Dissertation.

European Research Council. 2019. "Women in science - gender balancing at the ERC." ERC: european research council. Website. Retrieved October 9, 2021 (https://erc.europa.eu/news-events/magazine/woman-science-gender-balancing-erc).

Forscher, P.S., Cox, W.T.L., Brauer, M., Devine, P.G., 2019. Little race or gender bias in an experiment of initial review of NIH R01 grant proposals. Nat. Hum. Behav. 3 (3), 257–264.

Ginther, D.K., Hayes, K.J., 2003. Gender differences in salary and promotion for faculty in the humanities 1977–95. J. Hum. Resour. 38 (1), 34–73.

Ginther, D.K., Heggeness, M.L., 2020. Administrative discretion in scientific funding: evidence from a prestigious postdoctoral training program. Res. Policy 49 (4), 103953.

Goldin, C., Rouse, C., 2000. Orchestrating impartiality: the impact of" blind" auditions on female musicians. Am. Econ. Rev. 90 (4), 715–741.

Handley, I.M., Brown, E.R., Moss-Racusin, C.A., Smith, J.L., 2015. Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. Proc. Natl. Acad. Sci. 112 (43), 13201–13206.

Heggeness, M.L., Evans, L., Pohlhaus, J.R., Mills, S.L., 2016. Measuring diversity of the national institutes of health-funded workforce. Acad. Med. J. Assoc. Am. Med. Coll. 91 (8), 1164.

Hosek, S.D., 2005. Is There Gender Bias in Federal Grant Programs? Research Brief. RAND Corporation.

Jayasinghe, U.W., Marsh, H.W., Bond, N., 2003. A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. J. R. Stat. Soc. Ser. A 166 (3), 279–300 (Statistics in Society).

Kahneman, D., Miller, D.T., 1986. Norm theory: comparing reality to its alternatives. Psychol. Rev. 93 (2), 136.

Knobloch-Westerwick, S., Glynn, C.J., Huge, M., 2013. The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration Interest. Sci. Commun. 35 (5), 603–625.

Lamont, M., 2009. How Professors Think. Harvard University Press.

Lawson, C., Geuna, A., Finardi, U., 2021. The Funding-productivity-gender nexus in science, a multistage analysis. Res. Policy 50 (3), 104182.

Ley, T.J., Hamilton, B.H., 2008. The gender gap in NIH grant applications. Science 322 (5907), 1472–1474.

Marsh, H.W., Jayasinghe, U.W., Bond, N.W., 2008. Peer Reviews of Grant Applications: Effects of Applicant and Assessor Gender. UK: SELF Research Centre, University of Oxford, Oxford.

Marsh, H.W., Bornmann, L., Mutz, R., Daniel, H.D., O'Mara, A., 2009. Gender effects in the peer reviews of grant proposals: a comprehensive meta-analysis comparing traditional and multilevel approaches. Rev. Educ. Res. 79 (3), 1290–1326.

Milkman, K.L., Akinola, M., Chugh, D., 2015. What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. J. Appl. Psychol. 100 (6), 1678.

Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., Graham, M.J., Handelsman, J.O., 2012. Science faculty's subtle gender biases favor male students. Proc. Natl. Acad. Sci. 109 (41), 16474–16479.

Mutz, R., Bornmann, L., Daniel, H.D., 2012. Does gender matter in grant peer review? An empirical investigation using the example of the Austrian science fund. Z. Für Psychol. 220 (2), 121–129. https://doi.org/10.1027/2151-2604/a000103.

National Research Council, 2010. Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty. National Academies Press.

Nature Neuroscience, 2006. Women in Neuroscience. Nature Neuroscience 9, 853.

Niederle, M., Vesterlund, L., 2007. Do Women shy away from competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101. https://doi.org/10.1162/qjec.122.3.1067.

Nowlis, S.M., Simonson, I., 1997. Attribute–task compatibility as a determinant of consumer preference reversals. J. Market. Res. 34 (2), 205–218.

Oliveira, D.F.M., Ma, Y., Woodruff, T.K., Uzzi, B., 2019. Comparison of national institutes of health grant amounts to first-time male and female principal investigators. JAMA 321 (9), 898. https://doi.org/10.1001/jama.2018.21944.

Ridgeway, C.L., Correll, S.J., 2004. Unpacking the gender system: a theoretical perspective on gender beliefs and social relations. Gender Soc. 18 (4), 510–531.

Rivera, L.A., 2017. When two bodies are (not) a problem: gender and relationship status discrimination in academic hiring. Am. Sociol. Rev. 82 (6), 1111–1138.

Rivera, L.A., Tilcsik, A., 2019. Scaling down inequality: rating scales, gender bias, and the architecture of evaluation. Am. Sociol. Rev., 0003122419833601

Roper, R.L., 2019. Does gender bias still affect women in science? Microbiol. Mol. Biol. Rev. 83 (3), e00018–e00019.

Rossiter, M.W., 1993. The Matthew Matilda effect in science. Soc. Stud. Sci. 23 (2), 325–341.

Sandström, U., Hällsten, M., 2007. Persistent nepotism in peer-review. Scientometrics 74 (2), 175–189.

Sheltzer, J.M., Smith, J.C., 2014. Elite male faculty in the life sciences employ fewer women. Proc. Natl. Acad. Sci. 111 (28), 10107–10112.

Squazzoni, F., Bravo, G., Dondio, P., Farjam, M., Marusic, A., Menhami, B., Willis, M., Birukou, A., 2020. No evidence of any systematic bias against manuscripts by women in the peer review process of 145 scholarly journals. Working paper. https://doi.org/10.31235/osf.io/gh4rv.

Storage, D., Horne, Z., Cimpian, A., Leslie, S.J., 2016. The frequency of 'Brilliant' and 'Genius' in teaching evaluations predicts the representation of women and African Americans across fields. PLoS ONE 11 (3), e0150194.

Tregenza, T., 2002. Gender bias in the refereeing process? Trends Ecol. Evol. 17 (8), 349–350 (Amst.).

U.S. Government Accountability Office, 2015. Women in STEM Research: Better Data and Information Sharing Could Improve Oversight of Federal Grant-Making and Title IX Compliance. U.S. Government Accountability Office.

Van der Lee, R., Ellemers, N., 2015. Gender contributes to personal research funding success in the Netherlands. Proc. Natl. Acad. Sci. 112 (40), 12349–12353. https://doi.org/10.1073/pnas.1510159112.

Volker, B., Steenbeek, W., 2015. No Evidence that gender contributes to personal research funding success in the Netherlands: a reaction to van Der Lee and Ellemers. Proc. Natl. Acad. Sci. 112 (51), E7036–E7037. https://doi.org/10.1073/pnas.1519046112.

Waisbren, S.E., Bowles, H., Hasan, T., Zou, K.H., Emans, S.J, Goldberg, C., Gould, S., Levine, D., Lieberman, E., Loeken, M., Longtine, J., Nadelson, C., Patenaude, A.F., Quinn, D., Randolph, A.G., Solet, J.O.M., Ullrich, N., Walensky, R., Weitzman, P., Christou, H., 2008. Gender differences in research grant applications and funding outcomes for medical school faculty. J. Women's Health 17 (2), 207–214. https://doi.org/10.1089/jwh.2007.0412.

Ward, J.E., Donnelly, N., 1998. Is there gender bias in research fellowships awarded by the NHMRC? Med. J. Aust. 169 (11–12), 623–624.

Webb, T.J., O'Hara, B., Freckleton, R.P., 2008. Does Double-Blind Review Benefit Female Authors? Trends Ecol. Evol. 23 (7), 351–353 (Amst.).

Whittaker, R.J., 2008. Journal review and gender equality: a critical comment on Budden et al. Trends Ecol. Evol. 23 (9), 478–479 (Amst.).

Williams, WM., Ceci, SJ., 2015. National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. Proc. Natl. Acad. Sci. 112 (17), 5360–5365.

Witteman, H.O., Hendricks, M., Straus, S., Tannenbaum, C., 2019. Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. Lancet North Am. Ed. 393 (10171), 531–540. https://doi.org/10.1016/S0140-6736(18)32611-4.

Wold, A., Wennerás, C., 2010. Nepotism and Sexism in Peer-Review. Women, Science, and Technology. Routledge, pp. 64–70.