




ARTICLE




<https://doi.org/10.1057/s41599-022-01204-6>

OPEN

# Gender diversity of research consortia contributes to funding decisions in a multi-stage grant peer-review process

Stefano Bianchini<sup>1</sup>, Patrick Llerena<sup>1</sup>, Sila Öcalan-Özel<sup>1</sup> & Emre Özel<sup>1</sup>

This study seeks to draw connections between the grant proposal peer-review and the gender representation in research consortia. We examined the implementation of a multi-disciplinary, pan-European funding scheme—EUROpean COllaborative REsearch Scheme (2003–2015)—and the reviewers’ materials that this generated. EUROCORES promoted investigator-driven, multinational collaborative research in multiple scientific areas and brought together 9158 Principal Investigators (PI) who teamed up in 1347 international consortia that were sequentially evaluated by 467 expert panel members and 1862 external reviewers. We found systematically unfavourable evaluations for consortia with a higher proportion of female PIs. This gender effect was evident in the evaluation outcomes of both panel members and reviewers: applications from consortia with a higher share of female scientists were less successful in panel selection and received lower scores from external reviewers. Interestingly, we found a systematic discrepancy between the evaluative language of written review reports and the scores assigned by reviewers that works against consortia with a higher share of female participants. Reviewers did not perceive female scientists as being less competent in their comments, but they were negatively sensitive to a high female ratio within a consortium when scoring the proposed research project.

<sup>1</sup>BETA, CNRS, Université de Strasbourg, 67000 Strasbourg, France. email: [s.bianchini@unistra.fr](mailto:s.bianchini@unistra.fr)

## Introduction

Gender is one of the strongest social constructs employed for stereotyping others (Wood and Eagly, 2012; Haines et al., 2016). When men succeed in a domain that is culturally defined as masculine, their success is attributed to innate ability and talent, while women's success is typically attributed to external factors, including luck (Swim and Sanna, 1996). Words such as *intelligent* and *competent* are found within the cluster of positive traits of men (Abele and Wojciszke, 2007) but not of women (Eagly and Karau, 2002; Carli et al., 2016). As a result, women are generally accorded less status and power than men (Harris, 1991; Diekmann and Eagly, 2000; Fiske et al., 2002).

Academia is no exception to this behaviour. A number of studies point out that women are more often profiled as being less capable and skilled in the sciences (Foschi et al., 1994; Moss-Racusin et al., 2012; Chubb and Derrick, 2020), and are, thereby, devalued as scientific leaders (Ellemers et al., 2012). This, in turn, can negatively affect peer evaluations of women's scientific merit and success, systematically downgrading their qualifications and amplifying disparities in access to the resources—primarily, the research funding—needed to conduct science (Bedi et al., 2012; Bornmann et al., 2007; Alvarez et al., 2019; Huang et al., 2020). Such obstacles to research funding can reduce the presence of women in positions of scientific leadership and have important secondary effects on their mobility and career advancement (Bloch et al., 2014; Geuna, 2015).

Scholars refer to this phenomenon as *gender bias in grant allocation*. The term 'bias' is generally used to describe the representations that are produced by sub-optimal decision-making based on systematic simplifications and deviations from the tenets of rationality (Cosmides and Tooby, 1994; Haselton et al., 2016; Korteling et al., 2017). As such, an individual may attribute certain attitudes and stereotypes to another person based on observable characteristics, and in this case, gender. Previous research on gender bias in the peer-review of grant applications provides mixed evidence and is somewhat inconclusive. Some studies found a gender bias (Wenneras and Wold, 1997; Viner et al., 2004; Husu and De Cheveigné, 2010); others did not (Bazeley, 1998; Ley and Hamilton, 2008; Ceci and Williams, 2011; Lawson et al., 2021). The existing studies generally focus on a single stage in the evaluation (often those involving panel experts or external reviewers); on applications made at the individual level; and on a single country—see Table 1 for details.

Here we examined the implementation of a multi-disciplinary, pan-European funding scheme—EUROpean COLlaborative REsearch (EUROCORES) Scheme—and the reviewers' materials that this generated. Our research contributes to the debate on "gender in science" by investigating the associations between the proportion of female PIs within a consortium and project evaluation, through *all* stages of peer review, and for *all* scientific domains between 2003 and 2015.

Our study has three main objectives. First, we examined whether a potential gender bias is directed also against groups of individuals (i.e., research consortia), and not only against the individual PI, as revealed in previous studies. Second, EUROCORES data allowed us to examine whether this gender effect is evident in the decision-making process of groups of individuals (expert panels) and single evaluators (external reviewers), and how the effect propagates through the subsequent stages of the evaluation process. Third, we provided a lexicon-based sentiment analysis of the written reports of external reviewers to examine whether the sentiment polarity and the rate of emotion in the review texts are consistent with the review scores.

Our study contributes to explain the persistent underrepresentation of women in top academic roles, with important implications for institutions and policy makers.

## Data and methods

**The EUROCORES scheme.** The data for our analyses are drawn from the multi-stage peer-review evaluation process of the EUROCORES scheme (2003–2015). The aim of this scheme was to promote cooperation between national funding agencies in Europe by providing a mechanism for the collaborative funding of research on selected priority topics in and across all scientific domains. EUROCORES was based on a number of Research Programmes, the topics of which were selected through an annual call for themes, and which, in turn, comprised a number of Collaborative Research Projects (CRPs). Each CRP included at least three Individual Projects (IPs), each led by a Principal Investigator (PI) affiliated with a European university. CRP consortia worked together on a common work-plan and towards the common goals set out in their Outline Proposal (OP) and, subsequently, in their Final Proposal (FP). EUROCORES, which was terminated in 2015, supported 47 Research Programmes in different areas of research with an overall budget of some 150M euros.

The evaluation process of the EUROCORES scheme consisted of three consecutive stages (Fig. 1). At the first stage, the Expert Panels were responsible for evaluating Outline Proposals prior to inviting a consortium to submit a Final Proposal. Evaluations took place in separate face-to-face meetings over 1–2 days. Panels were made up, on average, of 12 members and each member was a spokesperson for three proposals (i.e., for at least nine Individual Projects). Panel decisions were made through consensus building, a process involving the interaction of peers within a scientific group. At the second stage, Final Proposals were received from selected applications and were sent for written reports to (at least) three anonymous referees. Reviewers were asked to complete a standardized evaluation form comprising 8–10 sections, each focusing on different criteria such as scientific quality, project feasibility, team interdisciplinarity, and others. Typically, the referee had 4 weeks to provide a written assessment of the proposal. For each section, a score was assigned on a 5-point Likert scale and comments were made in a dedicated space (at least 100 words for each question). One reviewer was responsible for evaluating just one project on each occasion. Thus, the overall evaluation was the result of an individual decision-making process with no (particularly strong) time constraints. At the third stage, once all the review reports had been received, expert panel members (the same as in the first stage of the evaluation) met a second time to make a decision based on the application, the referees' comments, the replies from applicants and open discussion. At this stage, panel members selected applications to be recommended for funding. The overall success rate for EUROCORES grant applications was ~13%.

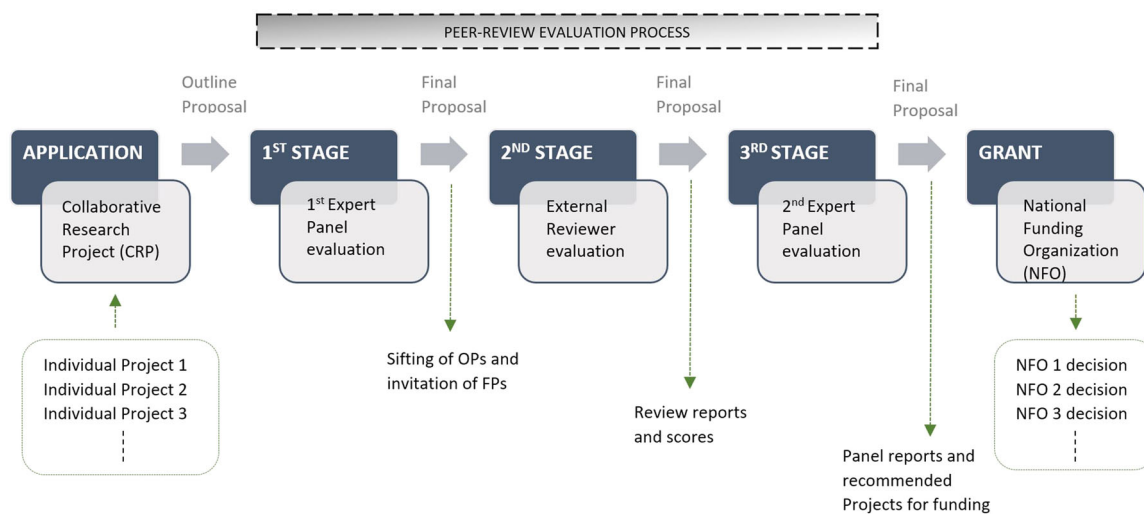
In this study, we examined the effect of the gender composition of the research consortia on grant evaluation decisions at three stages of the review process: *first expert panel evaluation* (1st stage); *external reviewer evaluation* (2nd stage); and *second expert panel evaluation* (3rd stage).

**The sample.** Our raw data contained information on 10,533 applicants, who teamed up to submit 1642 Outline Proposals [886 accepted and 756 rejected—success rate: 53%] and 886 Final Proposals [223 accepted and 663 rejected—success rate: 25%], and on 2182 external reviewers and 491 panel members throughout the three stages of the evaluation process. For each CRP application, which is our unit of observation, the data contain information on name, year of birth, gender and institutional affiliation of the applicants for each PI and evaluators (panel members and external reviewers), the application dates,

**Table 1 Previous empirical studies on gender bias in grant peer-review.**

Reference	Country	RFO	Period	Discipline	Dependent Variable	Gender effect
Banal-Estañol et al. (2019)	UK	EPSRC	1991-2007	Engineering and Physical Sciences	Awarded or not	No
Bautista-Puig et al. (2019)	EU	ERC	2007-2016	All Disciplines	Awarded or not	Yes
Bedi et al. (2012)	UK	Wellcome Trust	2000-2008	Biomedical Sciences	Awarded or not	Yes
Bol et al. (2022)	Netherlands	NWO	2005-2016	All Disciplines	Scores + Evaluation stages	Yes
Burns et al. (2019)	Canada	CIHR	2000-2014	Biomedical Sciences	Awarded or not	Yes
Cañibano et al. (2009)	Spain	Ramón y Cajal Programme	2005-2006	Physics, Molecular Biology, Philosophy and Philology	Awarded or not	No
Ginther et al. (2011)	USA	NIH	2000-2006	Biomedical Sciences	Awarded or not	Yes
Ginther et al. (2016)	USA	NIH	2000-2006	Biomedical Sciences	Awarded or not	Yes
Head et al. (2013)	UK	Various	1997-2010	Biomedical Sciences	Awarded or not	Yes
Jagsi et al. (2009)	USA	NIH	1993-2007	Biomedical Sciences	Awarded or not	Yes
Lerchenmueller and Sorenson (2018)	USA	NIH	1985-2015	Biomedical Sciences	Awarded or not	Yes
Magua et al. (2017)	USA	NIH	2010-2014	Biomedical Sciences	Scores + Sentiment analysis	Yes
Mutz et al. (2014)	Austria	ASF	1999-2009	All Disciplines	Scores + Awarded or not	Small
Pohlhaus et al. (2011)	USA	NIH	1999-2004	Biomedical Sciences	Awarded or not	Mixed Results
Sandström and Hällsten (2008)	Sweden	VR	2004	Biomedical Sciences	Scores	No
Severin et al. (2020)	Switzerland	SNSF	2009-2016	All Disciplines	Scores	Yes
Tamblyn et al. (2018)	Canada	CIHR	2001-2005	Biomedical Sciences	Awarded or not	Yes
van der Lee and Ellemers (2015)	Netherlands	NWO	2010-2012	All Disciplines	Scores + Sentiment analysis + Evaluation stages	Yes
Witteman et al. (2019)	Canada	CIHR	2011-2016	Biomedical Sciences	Awarded or not	Yes
Yip et al. (2020)	Hong Kong	Univ. of Hong Kong	2015-2020	Social Sciences	Awarded or not	No

Gender bias in research funding has been studied by different research areas, across disciplines and countries for more than three decades. See, e.g., Bornmann et al. (2007) and Marsh et al. (2009) for a systematic review and meta-analysis based on the literature published before 2007. The studies mentioned in this table start from 2008. Only a few studies examine the review materials and each step of the grant evaluation process.

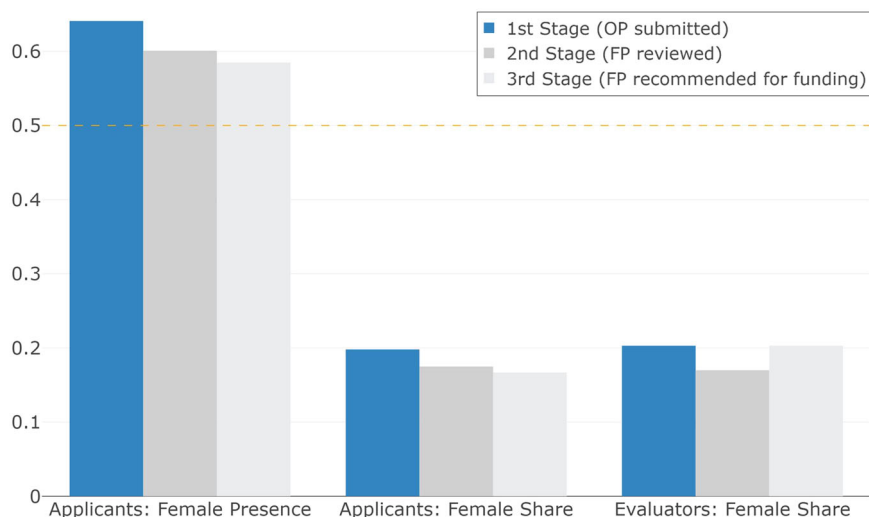


**Fig. 1 EUROCORES multi-stage evaluation process.** Submission of Outline Proposals (OPs) of Collaborative Research Projects (CRPs) corresponds to the APPLICATION phase. In the 1st STAGE of peer-review evaluation, OPs are sifted by Expert Panels and selected projects are invited to submit Final Proposals (FPs). In the 2nd STAGE, FPs are sent to at least 3 different External Reviewers to get written reports and scores for each FP. In the 3rd STAGE, Expert Panels select the FPs to be recommended for funding based on the reviewers' reports and scientific merit. In the GRANT Phase, each Individual Project (IP) receives a funding decision from its National Funding Organization (NFO).

the amount of funding requested as well as the review reports and scores given by the evaluators at corresponding stages.

The original data required some pre-processing. First, about 20% of the original observations were incomplete—e.g., they were missing the gender of the applicant, his or her age, and affiliation. Thus, where possible, we manually retrieved the missing information from

each researcher's personal and/or institutional web page. Second, the sample for the analysis was limited to proposals with complete information for all consecutive stages. Hence, the final sample consists of 1347 CRPs from 9158 unique applicants [sampling fraction: 82% and 87%, respectively]; 467 individual panel members [95%]; and 1862 written reports from external reviewers [85%].



**Fig. 2 Gender statistics for applicants and evaluators.** ‘Female Presence’ represents the proportion of consortia with at least one female PI. ‘Female Share’ is the proportion of female PIs in consortia or evaluator groups.

Project selection occurred in the first and third phases. The success rate in the first phase (OP to FP) was 38% ( $n = 511$  projects;  $n = 3579$  applicants), and in the third phase about 60% ( $n = 306$  projects;  $n = 2200$  applicants). A cursory glance at gender statistics suggests that female participation declined in each consecutive stage: 19.8% in the outline proposals, 17.5% in the final proposals and 16.7% after the second expert panel evaluation. The share of female evaluators is roughly 20% (Fig. 2).

**Statistical analysis.** We explored the association between the proportion of female PIs in EUROCORES research consortia and the decisions made by different evaluators throughout the three stages of the evaluation process.

The decisions of the panel experts were binary (i.e., selected vs. not selected) while the external reviewers assigned scores on a 5-point Likert scale and justified these scores in a short written submission (reviewer’s report). Thus, we used probit (for the first and third stages) and ordinary least-squares (for the second stage) models with standard errors clustered at the research programme level. The dependent variable in the second stage of the peer-review process is only observable for a portion of the data—i.e., those proposals that passed the first stage. The sample selection bias that might result from the sequentiality of the evaluation process was rectified by Heckman’s two-step estimation procedure (Heckman, 1979; Puhani, 2000).

For the first and third stages, we are interested in the factors—*in primis* the gender composition of a consortium—that influence the likelihood of being selected by panel experts. For the second stage of selection (external reviewers), we modelled both the reviewers’ scores and the sentiments associated with their reviews, and test for inconsistencies between scores and sentiments.

More specifically, we determined sentiment polarity in the reviewers’ reports using the VADER—Valence Aware Dictionary and sEntiment Reasoner—algorithm (Hutto and Gilbert, 2014) and developed a list of evaluative terms for both project and applicants using the Word2vec model (Mikolov et al., 2013). The general sentiment analysis tool VADER captured the emotional polarity (positive/negative) and intensity (strength) of reviews in relation to the applicants and their research proposals. Alternative algorithms were also tested (i.e., Syuzhet and sentimentR).

The evaluative terms were obtained through a Word2vec Skip-Gram model (Mikolov et al., 2013), trained on the corpus of reviewers’ reports. Technically, the words of a vocabulary of size  $V$

were positioned in a  $D$ -dimensional space. Hence, each word was represented by a  $D$ -dimensional continuous vector—i.e., the word representation. Words that tend to appear close to each other in the corpus have similar vector representations. This property allowed us to identify a list of lexical features (bi-grams) that co-occurred with high probability in a context window surrounding three terms: ‘principal investigator’, ‘consortium’ and ‘team’.<sup>1</sup> The identified bi-grams refer to the adjectives and adverbs used by reviewers to judge PIs and consortia, and include evaluative terms such as ‘internationally recognized’, ‘highly qualified’, or ‘world-renowned’. These terms are similar to those identified in previous studies on textual analysis of reviewers’ reports (see, e.g., Magua et al., 2017). We arbitrarily selected the most frequent 30 attributes and created a binary indicator taking a value of 1 if a review included at least one from the list, 0 otherwise.<sup>2</sup>

All models included fixed effects for the year, scientific domains, and research programmes, and a comprehensive set of covariates. We retrieved individual data from the Web of Science (WoS) and constructed several consortium-wide bibliometric indicators, including scientific productivity (total number of publications of the consortium normalized by the consortium size, in logs), participation of highly cited scientists, measures of research diversity and interdisciplinarity (Blau index), cognitive proximity (share of overlapping WoS subject categories between consortium and evaluators) and network proximity (if the consortium and evaluators have at least one common co-author prior to the grant application).

We also considered other factors that might influence evaluation decisions: the seniority of the consortium (average age of consortium members, in logs), its size and institutional reputation (whether the consortium includes at least one member affiliated with a university belonging to the Top 100 Shanghai Ranking), partnerships with the private sector (whether the consortium includes at least one member with a private sector affiliation), the size of the budget requested, experience with previous EUROCORES grants (whether the consortium includes at least one member having a EUROCORES-granted project prior to application), and the number of participating countries within a consortium. Finally, we included some characteristics of the evaluators, including gender, scientific productivity, age, and institutional reputation. We also took into account panel workload and the number of panel members.

Full details on variable construction, definitions, and descriptive statistics can be found in the SI material (Tables S1 and S2).

**Table 2 Coefficient estimates—First expert panel evaluation.**

	(1)	(2)	(3)
Female ratio	−0.197** (0.079)	−0.179** (0.076)	−0.213*** (0.081)
Gender balanced	0.011 (0.058)	0.001 (0.062)	0.013 (0.061)
Productivity	0.040 (0.030)	0.040 (0.026)	0.022 (0.028)
Star scientist	0.113* (0.058)	0.056 (0.055)	0.094* (0.053)
Cognitive proximity	0.118** (0.046)	0.064 (0.043)	0.135** (0.054)
Research diversity	−0.002 (0.048)	−0.092* (0.052)	−0.123*** (0.039)
Network proximity	−0.026 (0.101)	−0.067 (0.098)	−0.211** (0.084)
Size		0.124*** (0.040)	0.079* (0.047)
Age		−0.163 (0.173)	−0.208 (0.172)
Shanghai ranking		0.080** (0.033)	0.097*** (0.032)
Private partnership		−0.085 (0.072)	−0.090 (0.071)
Past EUROCORES grant		0.127** (0.052)	0.154*** (0.050)
Budget requested		0.053** (0.023)	0.027 (0.021)
Number of countries		0.156*** (0.037)	0.115** (0.045)
Female ratio panel			0.023 (0.198)
Productivity panel			0.056 (0.056)
Panel size			−0.154*** (0.059)
Age panel			0.339 (0.389)
Panel workload			−0.216*** (0.049)
Shanghai ranking panel			−0.030 (0.186)
Programme fixed effects	✓	✓	
Year dummies	✓	✓	✓
Domain dummies	✓	✓	✓
Observations	1347	1347	1347
AIC	1532.531	1468.846	1511.633
BIC	1777.196	1749.950	1673.008
Log likelihood	−719.265	−680.423	−724.817
Deviance	1438.531	1360.846	1449.633

This table reports marginal probability effects of the probit estimation, while holding all regressors at their mean value. The dependent variable is 'Passed 1st Stage' as defined in Table S1. All other explanatory variables are also defined in Table S1. The three models include different configurations of explanatory variables to test the sensitivity of the estimation. Model (3) is used as the first step in Heckman's selection model. Robust standard errors in parentheses, clustered at programme-level: \*\*\*, \*\*, \* indicate significance at the 1%, 5% and 10% level, respectively.

**Results**

**First stage: expert panels select OPs.** At this stage, we modelled the probability for an outline proposal to pass the first panel selection. Our data provide compelling evidence that consortia with a higher proportion of female PIs were at a comparative disadvantage in this first evaluation [Estimate: −0.213—std. error: 0.081—*p*-value < 0.01]. The most complete model specification, Table 2—Column 3, suggests that a 1% increase in the share of

female scientists significantly reduced the likelihood of advancing to the second stage by ~0.2%.

Other factors played a significant role in selection. For example, consortium size and the number of different participating countries were important determinants of success. Cognitive proximity between applicants and panel experts, along with prior success with EUROCORES applications, also had a positive influence on panel decisions. In contrast, consortia with a diverse research background and in close proximity to the panel members' collaboration network were more likely to find their application rejected. Apart from panel size and workload, the other panel characteristics had no significant impact on their evaluations.

**Second stage: external reviewers assess FPs.** At this stage, full proposals were sent for external review to at least three anonymous referees. We considered the average of the scores assigned to the scientific quality of the proposed research and the qualifications of the PIs as a proxy for scientific merit. This approach is justified as the format of the two questions related to the scientific quality of the proposal and the qualifications of the PIs remained unchanged throughout the period, while the format of the other questions underwent some minor changes, making them unsuitable for statistical analysis. We constructed reviewer sentiment measures and drew up a list of evaluative terms for the corpus of reviewers' reports, and explored the extent to which scores and language patterns in the reports differed for consortia with a high proportion of female scientists.

First, we found a negative relationship between gender and the reviewers' scores. Our data, Table 3—Column 6, show that a 1% increase in the proportion of female PIs within a consortium resulted in a 0.356% fall in the scores received [Estimate: −0.356—std. error: 0.175—*p*-value < 0.05]. The data also confirm that teams with greater scientific productivity scored higher, and evaluators penalized consortia closer to their areas of expertise, in line with previous research (Boudreau et al., 2016).

Second, we found a discrepancy between the scores and the written assessments contained in the reviewers' reports. Indeed, the analysis showed that the valence scores of the review corpus were neither positively nor negatively affected by the gender composition of the consortia (Table 4) [VADER estimate: 0.024—std. error: 0.071—*p*-value > 0.10; Presence of evaluative terms: 0.042—std. error: 0.091—*p*-value > 0.10]. Hence, sentiment scores as well as the presence of evaluative terms were largely unrelated to review scores. Reviewers did not perceive female PIs as being less competent in their comments; however, they were negatively sensitive to a high female ratio within a consortium when scoring the proposed research project. These discrepancies between the quantitative (scores) and qualitative (sentiments and evaluative terms) aspects of the reviewers' reports imply that even though evaluation language seems to be similar, consortia with higher share of female PIs had significantly lower scores.

**Third stage: expert panels make funding recommendation.** At the third stage, we modelled the probability for a final proposal to receive a recommendation for funding.

As shown in Table 5, we found no direct evidence of gender composition of research consortia in panel decisions [Estimate: −0.089—standard error: 0.191—*p*-value > 0.10]. However, the decisions were strongly associated with the reviewers' scores, thus complying with EUROCORES guidelines that state the reviewers' reports should be considered as constituting the main basis for evaluations. Yet, we have already seen that the reviewers' scores seemed to be biased towards consortia with a higher proportion

**Table 3 Coefficient estimates—external reviewer evaluation (scores).**

	(1)	(2)	(3)	(4)	(5)	(6)
Female ratio	-0.402*** (0.144)	-0.284* (0.152)	-0.359** (0.143)	-0.339* (0.175)	-0.367** (0.143)	-0.356** (0.175)
Gender balanced	0.072 (0.083)	0.061 (0.084)	0.075 (0.083)	0.072 (0.084)	0.085 (0.084)	0.084 (0.084)
Productivity	0.213*** (0.040)	0.195*** (0.041)	0.211*** (0.041)	0.208*** (0.045)	0.213*** (0.041)	0.211*** (0.045)
Star scientist	-0.203*** (0.076)	-0.260*** (0.077)	-0.222*** (0.075)	-0.228*** (0.080)	-0.222*** (0.075)	-0.226*** (0.079)
Cognitive proximity	-0.162*** (0.060)	-0.180*** (0.060)	-0.174*** (0.060)	-0.175*** (0.060)	-0.137** (0.061)	-0.138** (0.062)
Research diversity	0.029 (0.083)	-0.015 (0.084)	-0.037 (0.089)	-0.034 (0.089)	-0.048 (0.088)	-0.046 (0.089)
Size			-0.016 (0.073)	-0.024 (0.082)	-0.021 (0.073)	-0.025 (0.082)
Age			0.213 (0.217)	0.229 (0.228)	0.185 (0.219)	0.193 (0.229)
Shanghai ranking			0.092* (0.047)	0.083 (0.062)	0.093** (0.047)	0.089 (0.061)
Private partnership			-0.072 (0.122)	-0.065 (0.124)	-0.079 (0.121)	-0.076 (0.123)
Past EUROCORES grant			0.103* (0.062)	0.092 (0.079)	0.100 (0.061)	0.094 (0.079)
Budget requested			-0.011 (0.036)	-0.013 (0.038)	-0.013 (0.036)	-0.014 (0.038)
Number of countries			0.137 (0.093)	0.126 (0.105)	0.141 (0.092)	0.135 (0.105)
Female reviewer					-0.039 (0.055)	-0.039 (0.055)
Productivity reviewer					-0.059** (0.023)	-0.059** (0.023)
Age reviewer					0.064 (0.098)	0.064 (0.098)
Shanghai ranking reviewer					-0.0004 (0.045)	-0.0005 (0.045)
Inverse Mills ratio		-0.313*** (0.107)		-0.054 (0.248)		-0.028 (0.248)
Programme fixed effects	✓		✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Domain dummies	✓	✓	✓	✓	✓	✓
Observations	1862	1862	1862	1862	1862	1862
R <sup>2</sup>	0.138	0.142	0.144	0.144	0.148	0.148
Adjusted R <sup>2</sup>	0.117	0.121	0.121	0.120	0.122	0.122
Residual std. error	0.837	0.835	0.835	0.835	0.834	0.834
F statistic	6.747***	6.819***	6.115***	5.993***	5.809***	5.700***

This table reports coefficients of the OLS estimation. The dependent variable is 'Reviewer Score' as defined in Table S1. All other explanatory variables are also defined in Table S1. The six models include different configurations of explanatory variables to test the sensitivity of the estimation. Columns (2), (4) and (6) report Heckman's two-step estimated coefficients and corresponding Inverse Mills Ratios. Robust standard errors in parentheses, clustered at programme-level: \*\*\*, \*\*, \*, indicate significance at the 1%, 5% and 10% level, respectively.

of female members, implying that expert panel decisions may also be indirectly biased towards these consortia.

Besides the gender dimension, the data also show an important positive role of network proximity consistent with 'old boy' network patterns (Rose, 1989; Travis and Collins, 1991), as identified in previous studies (Wenneras and Wold, 1997; Sandström and Hällsten, 2008).

**Discussion**

In this article, we examined an original dataset from EURO-CORES grants scheme and explored the factors affecting evaluation outcomes at each stage of the peer-review process. Our analysis reveals a noteworthy interaction between the gender of applicants and peer review outcomes, and provides compelling evidence of a strong negative impact on consortia with a higher representation of female scientists. This gender effect is present in the evaluation outcomes of both panels and external reviewers.

Our results also show that there is a mismatch between text and scores in external review reports that runs against consortia with a higher proportion of female PIs.

There is a growing body of theoretical and empirical research on gender bias. Biases in judgements are part of human nature and are often the results of some heuristics of thinking under uncertainty (Tversky and Kahneman, 1974). For example, the heuristic in evaluating a grant proposal may be to rely too heavily on easily perceived characteristics of applicants, which may be effective in part (think of scientific excellence), but prone to producing biases (Shafir and LeBoeuf, 2002). This does not necessarily mean that the bias is intentional; on the contrary, it may well occur outside of the decision maker's awareness (Kahneman, 2011). Our findings on the external reviewers' evaluations suggest that this might be the case. External reviewers may have had stereotypes that they did not report verbally, and that most likely occurred outside of their conscious awareness

**Table 4 Coefficient estimates—external reviewer evaluation (sentiments and evaluative terms).**

	SentimentR	Syuzhet	VADER	Evaluative terms
Female ratio	-0.035 (0.039)	0.456 (1.338)	0.024 (0.071)	0.042 (0.091)
Gender balanced	-0.002 (0.016)	1.333* (0.754)	-0.018 (0.034)	-0.022 (0.048)
Productivity	0.028** (0.011)	0.134 (0.425)	0.008 (0.017)	0.013 (0.026)
Star scientist	-0.012 (0.018)	-1.211* (0.730)	-0.058 (0.036)	0.056 (0.047)
Cognitive proximity	-0.005 (0.012)	0.064 (0.575)	0.004 (0.024)	0.021 (0.036)
Research diversity	-0.043** (0.019)	-0.614 (0.742)	0.038 (0.038)	-0.013 (0.048)
Size	0.053*** (0.020)	1.170* (0.659)	0.020 (0.033)	0.007 (0.046)
Age	0.072 (0.058)	1.416 (1.893)	0.043 (0.087)	0.055 (0.128)
Shanghai ranking	0.021 (0.014)	0.544 (0.516)	0.027 (0.026)	0.058* (0.035)
Private partnership	-0.035 (0.036)	-0.267 (1.102)	0.004 (0.051)	0.030 (0.080)
Past EUROCORES grant	0.028 (0.017)	0.763 (0.778)	0.002 (0.034)	0.009 (0.046)
Budget requested	-0.005 (0.009)	0.418 (0.328)	0.007 (0.015)	-0.001 (0.021)
Number of countries	-0.011 (0.021)	-1.112 (0.987)	-0.004 (0.046)	-0.036 (0.060)
Female reviewer	-0.013 (0.011)	0.848* (0.488)	0.001 (0.023)	0.022 (0.031)
Productivity reviewer	-0.009* (0.005)	-0.663*** (0.202)	-0.044*** (0.011)	0.003 (0.013)
Age reviewer	0.033 (0.020)	-2.696*** (0.872)	-0.127*** (0.041)	-0.038 (0.056)
Shanghai ranking reviewer	-0.026*** (0.010)	-0.473 (0.372)	-0.031 (0.020)	0.023 (0.026)
Inverse Mills ratio	0.067 (0.058)	-1.780 (2.028)	-0.007 (0.101)	0.030 (0.133)
Programme fixed effects	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓
Domain dummies	✓	✓	✓	✓
Observations	1862	1862	1862	1862
R <sup>2</sup>	0.081	0.119	0.136	0.062
Adjusted R <sup>2</sup>	0.053	0.092	0.109	0.033
Residual std. error	0.183	7.305	0.357	0.478
F statistic	2.896***	4.414***	5.157***	2.167***

This table reports coefficients of the OLS estimation [Columns(1)–(3)], and of the probit estimation [Column (4)]. The dependent variables are different sentiment scores obtained through different algorithms, the names of which appear in the column header. All other explanatory variables are defined in Table S1. Robust standard errors in parentheses, clustered at programme-level: \*\*\*, \*\*, \*, indicate significance at the 1%, 5% and 10% level, respectively.

and control, but they were, nevertheless, clearly manifest in their scores. In the literature, the term *implicit bias* is often used to refer to such attitudes and stereotypes in general (Mandelbaum, 2015; Frankish, 2016).

Overall, the findings reported here add to our understanding of gender bias in science by showing that such bias is not solely directed against the individual, as revealed in previous studies, but that it has a more pervasive effect and can involve groups of individuals too. The sequential nature of the grant evaluation process is designed to help screen proposals and applicants based on scientific merit, with subsequent steps functioning as ‘filters’ to keep the best proposals alive. However, our results show that precisely because of the sequential nature of the process, gender bias at one stage can indirectly influence decisions at subsequent stages.

Although our study emphasized a gender bias in the peer review process, we acknowledge that the effect found could be

caused by the concomitance of other factors. Four aspects seem particularly relevant to us. First, the text characteristics of the outline and full proposals. Indeed, we could not measure the quality of the written proposal nor the editing style that might be dependent on the gender composition of a consortium. Although gender bias is a more complex problem than just the differences between men and women in using language, writing style and word choice may have an important impact on grant evaluation and selection processes (Tse and Hyland, 2008; Kolev et al., 2020). Second, we did not have information on the applicants’ time allocation in work (e.g., teaching and administrative duties) and family (e.g., childcare and housework). Third, most EUROCORES projects involved interdisciplinary collaborations (e.g., physics and engineering, life and environmental sciences, biomedicine, social sciences and humanities), which made it impractical to investigate how gender bias varied by scientific macro-area. Finally, we cannot exclude the

**Table 5 Coefficient estimates—second expert panel evaluation.**

	(1)	(2)	(3)
Female ratio	0.097 (0.197)	0.010 (0.181)	−0.089 (0.191)
Gender balanced	0.019 (0.072)	0.033 (0.075)	0.077 (0.093)
Reviewers scores mean	0.685*** (0.109)	0.693*** (0.114)	0.708*** (0.106)
Reviewers scores variance	0.008 (0.043)	−0.000 (0.047)	0.039 (0.043)
Productivity	−0.020 (0.058)	−0.043 (0.053)	−0.046 (0.040)
Star scientist	0.114 (0.096)	0.136* (0.080)	0.151* (0.087)
Cognitive proximity	−0.012 (0.126)	0.023 (0.106)	−0.025 (0.112)
Research diversity	0.133 (0.164)	0.104 (0.150)	0.208 (0.151)
Network proximity	0.267*** (0.041)	0.252*** (0.037)	0.338*** (0.051)
Size		−0.014 (0.082)	−0.088 (0.083)
Age		−0.488 (0.382)	−0.262 (0.377)
Shanghai ranking		−0.001 (0.104)	−0.023 (0.097)
Private partnership		−0.470*** (0.163)	−0.185 (0.195)
Past EUROCORES grant		0.074 (0.061)	0.049 (0.069)
Budget requested		0.113*** (0.043)	0.078* (0.041)
Number of countries		0.013 (0.096)	0.056 (0.113)
Female ratio panel			−0.298 (0.294)
Productivity panel			0.005 (0.081)
Panel size			−0.353*** (0.100)
Age panel			−1.094** (0.555)
Panel workload			−0.281*** (0.071)
Shanghai ranking panel			0.114 (0.113)
Programme fixed effects	✓	✓	
Year dummies	✓	✓	✓
Domain dummies	✓	✓	✓
Observations	511	511	511
AIC	509.903	508.441	518.656
BIC	708.921	737.100	658.392
Log likelihood	−207.952	−200.221	−226.328
Deviance	415.903	400.441	452.656

This table reports marginal probability effects of the probit estimation, while holding all regressors at their mean value. The dependent variable is 'Passed 3rd Stage' as defined in Table S1. All other explanatory variables are also defined in Table S1. The three models include different configurations of explanatory variables to test the sensitivity of the estimation. Robust standard errors in parentheses, clustered at programme-level: \*\*\*, \*\*, \*, indicate significance at the 1%, 5% and 10% level, respectively.

existence of a self-selection bias in the decision of a female scientist to apply for a EUROCORES grant. Future research should address these limitations to achieve a better understanding of the gendered nature of evaluation in peer-review processes.

The EUROCORES Scheme was ended in December 2015 after almost 12 years of activity. However, the lessons learned from this scheme are still relevant today, as there are many similar national and international funding schemes managed by different Research Funding Organizations (RFOs) such as the European Research Council (ERC), the National Institute of Health (NIH), the National Science Foundation (NSF), the French National Research Agency (ANR), and the German Research Foundation (DFG), among others. These RFOs develop and implement assessment procedures similar to EUROCORES that mostly rely on external peer-review and expert panels to determine successful applicants. Our results are, therefore, relevant to policy makers and RFOs.

Clearly, we endorse calls for a more equitable grant peer review system in order to avoid all forms of conflicts of interest, including cognitive, social and other forms of proximity. Anonymizing the applicant's profile can only be a partial solution to the problem, and one that may not be particularly effective as the identity of the applicants must be known to the evaluator in order to access their bibliometric indicators, for instance their publication and citation records. A more drastic and, perhaps, more effective approach would be to actually inform panel members and reviewers ex-post about any biases observed in their decisions. If the bias exists, then it is there; there is little that can be done to rectify it. Yet, the taking of such an approach might go some way to increasing an evaluator's awareness and to changing their future behaviour.

**Data availability**

We obtained EUROCORES grant records from the European Science Foundation (ESF). EUROCORES data are not publicly available. Codes for data preparation, variable construction, statistical analysis and details on data access are available from the corresponding author upon request and/or on [GitHub](#).

Received: 26 May 2021; Accepted: 16 May 2022; Published online: 10 June 2022

**Notes**

- Some technical details about the estimation of word representations are in order. First, we removed all reviewers' reports with <15 words, a pre-defined set of stop words, and all words occurring <5 times in the corpus. Second, we pasted uni-grams into bi-grams depending on their co-occurrence (threshold equal to 50). Third, we used negative sampling. Hence, we estimated a logit model where the binary dependent variable indicates whether or not two terms are close in the corpus, at distance *c*. For each observed neighbouring term pair (success), one should add *k* 'negative samples' (failures). The results presented along this article were obtained with the following parameter settings. We set the dimensionality of the dense word representation to 512 dimensions (we tried other dimensions: 256, 300, 512 and 1024). We defined a context window (distance *c*) of 7 words from both sides around the target. For each observed neighbouring term pair, we draw *k* = 15 negative example.
- Most frequent 30 evaluative terms in reviewers' reports: connected internationally; considerable experience; excellent track; highly qualified; highly respected; international connection; international reputation; international standing; internationally competitive; internationally connected; internationally recognized; internationally renowned; leading expert; leading scientists; numerous cooperation; significant contribution; track record; world leader; world leading; world renowned.

**References**

Abele A, Wojciszke B (2007) Agency and communion from the perspective of self versus others. *J Personal Soc Psychol* 93(5):751–763

Alvarez SNE, Jagsi R, Abbuhl S, Lee C, Myers E (2019) Promoting gender equity in grant making: what can a funder do? *The Lancet* 393:e9–e11

Banal-Estañol A, Macho-Stadler I, Pérez-Castrillo D (2019) Evaluation in research funding agencies: are structurally diverse teams biased against? *Res Policy* 48. <https://doi.org/10.1016/j.respol.2019.04.008>



- Bautista Puig N, García-Zorita C, Mauleón E (2019) European research council: excellence and leadership over time from a gender perspective *Res Eval* 28:370–382. <https://doi.org/10.1093/reseval/rvz023>
- Bazeley P (1998) Peer review and panel decisions in the assessment of Australian research council project grant applicants: what counts in a highly competitive context? *High Educ* 35:435–452
- Bedi G, Dam NT, Munafo M (2012) Gender inequality in awarded research grants. *The Lancet* 380:474
- Bloch C, Graversen E, Pedersen HS (2014) Competitive research grants and their impact on career performance. *Minerva* 52:77–96
- Bol T, de Vaan M, van de Rijt A (2022) Gender-equal funding rates conceal unequal evaluations *Res Policy* 51(1):104399
- Bornmann L, Mutz R, Daniel H-D (2007) Gender differences in grant peer review: a meta-analysis. *J Informetr* 1:226–238
- Boudreau K, Guinan E, Lakhani K, Riedl C (2016) Looking across and looking beyond the knowledge frontier: intellectual distance, novelty, and resource allocation in science. *Manag Sci* 62:2765–2783
- Burns KEA, Straus SE, Liu K, Rizvi L, Guyatt G (2019) Gender differences in grant and personnel award funding rates at the Canadian Institutes of Health Research based on research content area: a retrospective analysis. *PLoS Med* 16(10):e1002935. <https://doi.org/10.1371/journal.pmed.1002935>
- Cañibano C, Otamendi J, Andújar I (2009) An assessment of selection processes among candidates for public research grants: the case of the Ramón y Cajal programme in Spain *Res Eval* 18:153–161
- Carli LL, Alawa L, Lee Y, Zhao B, Kim E (2016) Stereotypes about gender and science. *Psychol Women Q* 40:244–260
- Ceci SJ, Williams WM (2011) Understanding current causes of women's underrepresentation in science *Proc Natl Acad Sci USA* 108:3157–3162
- Chubb J, Derrick G (2020) The impact a-gender: gendered orientations towards research impact and its evaluation. *Palgrave Commun* 6:1–11
- Cosmides L, Tooby J (1994) Better than rational: evolutionary psychology and the invisible hand. *Am Econ Rev* 84:327–332
- Diekmann A, Eagly A (2000) Stereotypes as dynamic constructs: women and men of the past, present, and future. *Personal Soc Psychol Bull* 26:1171–1188
- Eagly A, Karau S (2002) Role congruity theory of prejudice toward female leaders. *Psychol Rev* 109(3):573–598
- Ellemers N, Rink F, Derks B, Ryan M (2012) Women in high places: when and why promoting women into top positions can harm them individually or as a group (and how to prevent this). *Res Organ Behav* 32:163–187
- Fiske S, Cuddy AJC, Glick P, Xu J (2002) A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J Personal Soc Psychol* 82 6:878–902
- Foschi M, Lai L, Sigerson K (1994) Gender and double standards in the assessment of job applicants. *Soc Psychol Q* 57:326–339
- Frankish K (2016) Playing double: Implicit bias, dual levels, and self-control. In: Brownstein M, Saul J (eds) *Implicit bias and philosophy, vol 1: Metaphysics and epistemology*. Oxford University Press.
- Geuna A (2015) *Global mobility of research scientists: the economics of who goes where and why*. Academic Press
- Ginther D, Schaffer W, Masimore B, Liu F, Haak L, Kington R (2011) Race, ethnicity, and NIH research awards *Science* 333:1015–1019. <https://doi.org/10.1126/science.1196783>
- Ginther D, Kahn S, Schaffer W (2016) Gender, race/ethnicity, and national institutes of health R01 research awards: is there evidence of a double bind for women of color? *Acad Med* 91:1. <https://doi.org/10.1097/ACM.0000000000001278>
- Haines E, Deaux K, Lofaro N (2016) ThD. Ginther, S. Kahn, and W. Schaffer. Gender, race/ethnicity, and national institutes of health R01 research awards: Is there evidence of a double bind for women of color? *Academic Medicine*, 91:1, 06 2016. doi: 10.1097/ACM.0000000000001278e times they are a-changing ... or are they not? A comparison of gender stereotypes, 1983–2014. *Psychol Women Q* 40:353–363
- Harris A (1991) Gender as contradiction *Psychoanal Dialogue* 1:197–224
- Haselton M, Nettle D, Andrews P (2016) The evolution of cognitive bias. In: Buss DM (ed) *Handbook of evolutionary psychology*. Wiley, New York, pp 968–987
- Head M, Fitchett J, Cooke M, Wurie F, Atun R (2013) Differences in research funding for women scientists: a systematic comparison of UK investments in global infectious disease research during 1997–2010 *BMJ Open* 3:e003362. <https://doi.org/10.1136/bmjopen-2013-003362>
- Heckman JJ (1979) Sample selection bias as a specification error *Econometrica* 47(1):153–161
- Huang J, Gates AJ, Sinatra R, Barabási A (2020) Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc Natl Acad Sci USA* 117:4609–4616
- Husu L, De Cheveigné S (2010) Gender and gatekeeping of excellence in research funding: European perspectives. In: Riegraf B, Aulenbacher B, Kirsch-Auwärter E, Müller U (eds) *Gender change in Academia: re-mapping the fields of work, knowledge, and politics from a gender perspective*. VS Verlag für Sozialwissenschaften, GWV Fachverlage GmbH, pp. 43–59
- Hutto C, Gilbert E (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the International AAAI conference on Web and social media*. Vol. 8, No. 1, pp 216–225
- Jagsi R, Motomura AR, Griffith K, Rangarajan S, Ubel P (2009) Sex differences in attainment of independent funding by career development awardees *Ann Intern Med* 151:804–811
- Kahneman D (2011) *Thinking, fast and slow*. Macmillan.
- Kolev J, Fuentes-Medel Y, Murray F (2020) Gender differences in scientific communication and their impact on grant funding decisions. *AEA Pap Proc* 110:245–49
- Korteling JE, Brouwer A-M, Toet A (2017) A neural network framework for cognitive bias. *Front Psychol* 9:1561
- Lawson C, Geuna A, Finardi U (2021) The funding-productivity-gender nexus in science, a multistage analysis. *Res Policy* 50:104182
- Lerchenmueller M, Sorenson O (2018) The gender gap in early career transitions in the life sciences *Res Policy* 47. 10.1016/j.respol.2018.02.009
- Ley T, Hamilton B (2008) The gender gap in NIH grant applications. *Science* 322:1472–1474
- Magua W et al. (2017) Are female applicants disadvantaged in national institutes of health peer review? combining algorithmic text mining and qualitative methods to detect evaluative differences in R01 reviewers' critiques. *J Women's Health* 26(5):560–570
- Mandelbaum E (2015) Attitude, inference, association: on the propositional structure of implicit bias. *Noûs* 50 3:629–658
- Marsh H, Bornmann L, Mutz R, Daniel H-D, O'Mara-Eves 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. <https://doi.org/10.3102/0034654309334143>
- Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. Preprint at arXiv preprint arXiv:1301.3781.
- Moss-Racusin C, Dovidio J, Brescoll V, Graham M, Handelsman J (2012) Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci USA* 109:16474–16479
- Mutz R, Bornmann L, Daniel H-D (2014) Testing for the fairness and predictive validity of research funding decisions: a multilevel multiple imputation for missing data approach using ex ante and ex-post peer evaluation data from the Austrian science fund *J Assoc Inf Sci Technol* 6. 10.1002/asi.23315
- Pohlhaus JR, Jiang H, Wagner RM, Schaffer WT, Pinn VW (2011) Sex differences in application, success, and funding rates for NIH extramural programs. *Acad Med* 86(6):759
- Puhani PA (2000) The Heckman correction for sample selection and its critique. *J Econ Surv* 14:53–68
- Rose S (1989) Women biologists and the "old boy" network. *Womens Stud Int Forum* 12:349–354
- Sandström U, Hällsten M (2008) Persistent nepotism in peer-review. *Scientometrics* 74:175–189
- Severin A, Martins J, Heyard R, Delavy F, Forstad A, Egger M (2020) Gender and other potential biases in peer review: cross-sectional analysis of 38250 external peer review reports. *BMJ Open*. 10:e035058. <https://doi.org/10.1136/bmjopen-2019-035058>
- Shafir E, LeBoeuf RA (2002) Rationality. *Annu Rev Psychol* 53:491–517
- Swim J, Sanna LJ (1996) He's skilled, she's lucky: a meta-analysis of observers' attributions for women's and men's successes and failures. *Personal Soc Psychol Bull* 22:507–519
- Tamblyn R, Girard N, Qian C, Hanley J (2018) Assessment of potential bias in research grant peer review in Canada *Can Med Assoc J* 190:E489–E499
- Travis G, Collins H (1991) New light on old boys: cognitive and institutional particularism in the peer review system. *Sci Technol Hum Values* 16:322–341
- Tse P, Hyland K (2008) Robot kung fu: gender and professional identity in biology and philosophy reviews. *J Pragmat* 40:1232–1248
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185:1124–1131
- van der Lee R, Ellemers N (2015) Gender contributes to personal research funding success in the Netherlands *Proc Natl Acad Sci USA* 112:12349–12353
- Viner N, Powell P, Green R (2004) Institutionalized biases in the award of research grants: a preliminary analysis revisiting the principle of accumulative advantage. *Res Policy* 33:443–454
- Wenneras C, Wold A (1997) Nepotism and sexism in peer-review. *Nature* 387:341–343
- Witteaman H, 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 *The Lancet* 393:531–540
- Wood W, Eagly A (2012) Biosocial construction of sex differences and similarities in behavior. *Adv Exp Soc Psychol* 46:55–123
- Yip PSF, Xiao Y, Wong CLH, Au TKF (2020) Is there gender bias in research grant success in social sciences?: Hong Kong as a case study *Humanit Soc Sci Commun* 7(1):1–10

### Acknowledgements

We are grateful to the members of the European Science Foundation (ESF) for their valuable support, especially Nicolas Walter who provided guidance in understanding the data and functioning of EUROCORES. We acknowledge the important feedback from Carolina Cañibano, Aldo Geuna, Cornelia Lawson, Francesco Lissoni, Michele Pezzoni, and Fabiana Visentin. This research was supported by the Agence Nationale de la Recherche (ANR)—AAPG Grant ANR-19-CE26-0014: GIGA (Gender Bias in Grant Allocation).

### Author contributions

SB, PL, SÖ, and EÖ designed research; EÖ developed the conceptual framework; SÖ and EÖ prepared and managed the data; SB performed statistical analysis; SB, PL, SÖ, and EÖ wrote the paper.

### Competing interests

The authors declare no competing interests.

### Ethical approval

Not applicable—this article does not contain any studies with human participants performed by any of the authors.

### Informed consent

Not applicable—this article does not contain any studies with human participants performed by any of the authors.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-022-01204-6>.

**Correspondence** and requests for materials should be addressed to Stefano Bianchini.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022