nature human behaviour

ARTICLES
https://doi.org/10.1038/s41562-019-0686-3

Committees with implicit biases promote fewer women when they do not believe gender bias exists

Isabelle Régner[1*], Catherine Thinus-Blanc[1], Agnès Netter[2], Toni Schmader[3,5] and Pascal Huguet[4,5*]

Whether gender bias contributes to women's under-representation in scientific fields is still controversial. Past research is limited by relying on explicit questionnaire ratings in mock-hiring scenarios, thereby ignoring the potential role of implicit gender bias in the real world. We examine the interactive effect of explicit and implicit gender biases on promotion decisions made by scientific evaluation committees representing the whole scientific spectrum in the course of an annual nationwide competition for elite research positions. Findings reveal that committees with strong implicit gender biases promoted fewer women at year 2 (when committees were not reminded of the study) relative to year 1 (when the study was announced) if those committees did not explicitly believe that external barriers hold women back. When committees believed that women face external barriers, implicit biases did not predict selecting more men over women. This finding highlights the importance of educating evaluative committees about gender biases.

Despite recent advances, women remain under-represented in the fields of science, technology, engineering and mathematics (STEM). This disparity is well documented[1–3]. Much more controversial is whether gender bias plays a role in women's under-representation in STEM. Experimental studies have examined the possible role of gender bias in contributing to women's under-representation, but have revealed conflicting patterns. Whereas several studies find a hiring preference for men over women[4–6], others have found that current recruitment strategies in the sciences instead benefit women over men[7,8]—a provocative finding that has been promoted as evidence that "Academic science isn't sexist"[9]. However, this debate mostly relies on hypothetical scenarios, introspective self-reports and questionnaire ratings at the expense of direct observations of high-stake decisions in the real world[10–12]. The only published study conducted with real hiring committees showed evidence that women are favoured as high school teachers in male-dominated subjects[8]. However, as pointed out by Stewart and Valian[12], there is little reason to expect that decisions made about high school teachers would generalize to decisions made for elite scientists seeking prestigious research positions.

A key shortcoming of past research is the failure to examine how and when variation in decision-makers' implicit associations and explicit gender beliefs predict hiring outcomes. In contrast with explicit beliefs that are conscious and deliberate, implicit associations are automatically activated and can lead to discriminatory responses that are independent of conscious intention[13,14]. Implicit associations that make it easier for people to connect science and mathematics with males rather than females can be measured by the implicit association test (IAT)[15–17]. The IAT is a widely used measure of implicit associations that have the potential to bias thought and behaviour[16,18]. The tendency to automatically associate males with science on the IAT is related not only to interest and performance in scientific domains at the individual level[19], but also to gender gaps in mathematics performance[17] and science participation[20] at the national level. However, it is not known whether—and perhaps more importantly when—this implicit science = male association predicts real-world hiring and tenure decisions in academic science. Research suggests that the effect of implicit associations on behaviour should be moderated by explicit beliefs and values[14].

Based on theory and research on attitude–behaviour processes[21,22], individuals' explicit beliefs can justify their implicit biases, amplifying the effects that implicit associations have on behaviour. For example, gender discrimination is increased when decision-makers who endorse stereotypes assume themselves to be objective and rational actors[23]. Alternatively, explicit beliefs can promote executive control to inhibit or counteract the effect of implicit associations on behaviour. Although these theories are typically applied to decision-making at the individual level, we assert that similar processes can unfold dynamically within groups charged with making collective decisions. That is, groups whose members explicitly reject gender discrimination as a problem and doubt women's ability to succeed might allow the implicit biases of those on the committee to inform the collective decisions they make. In contrast, groups that see systemic barriers to women's advancement as a problem to be addressed are more likely to have members who are motivated to suppress or counteract biases that might arise (either by themselves or others) during group discussion and decision-making[14].

We tested these hypotheses about gender biases and decision-making with existing evaluation committees (414 members overall) representing the whole scientific spectrum (from particle physics to political sciences) in the normal course of annual nationwide competitions for elite research positions in France. All candidates were accomplished research scientists who met the criteria of scientific excellence as defined by the governing body. The official mission of committees therefore is to go beyond easily quantifiable measures of productivity (the *h* index) to make subjective decisions that are also based on the perceived originality of candidates' scientific contributions among other qualitative parameters (otherwise, a calculator

[1]Laboratoire de Psychologie Cognitive, Aix Marseille Univ, CNRS, Marseille, France. [2]Institut National des Hautes Etudes de la Sécurité et de la Justice, CNRS, Paris, France. [3]University of British Columbia, Vancouver, British Columbia, Canada. [4]Laboratoire de Psychologie Sociale et Cognitive, Université Clermont Auvergne, CNRS, Clermont-Ferrand, France. [5]These authors contributed equally: Toni Schmader, Pascal Huguet.
*e-mail: isabelle.regner@univ-amu.fr; pascal.huguet@uca.fr

would be sufficient). The *h* index itself might be biased, since papers authored by women receive 10.4 ± 0.9% fewer citations than would be expected if the papers with the same non-gender-specific properties were written by men[24]. The evaluation of the candidates on both quantifiable and non-quantifiable parameters thus leaves room for the intrusion of an implicit science = male association to bias evaluations. Selection decisions for each field were made at the committee level, not by averaging individual decisions but first based on a group discussion for each candidate followed by a group discussion of the final list of selected candidates with the aim of reaching consensus. The outcomes of each committee's decisions had real implications for the careers of hundreds of highly accomplished research scientists, who if awarded a position would be added to the group of 4,759 elite researchers who held these posts at the time of the present study. Figure 1a illustrates the gender asymmetry within each academic discipline in the year before data collection began.

With approval from the governing body overseeing these competitions, we tested the relationship between measured committee-level gender biases (both implicit and explicit) and selection outcomes. More specifically, we examined the degree to which selection decisions favoured men over women (accounting for the ratio of men and women in the applicant pool) for committees whose members hold a stronger implicit science = male implicit association and a weaker explicit belief that women face external barriers such as discrimination that constrain their success. The governing body provided the research team with the final selection decisions from all 40 committees for each of two consecutive years (committee members were the same from year 1 to year 2). At year 1, immediately before the start of committee sessions, all committees were informed that the governing body had authorized a research study to examine whether committees' selection decisions could be biased against women. Immediately after committees started their work for year 1 selections, they were invited to complete the gender–science IAT[17] assessing implicit biases and a questionnaire measuring their explicit beliefs (see Fig. 1b for the timeline). In the questionnaire, committee members were asked to rate their attributions for current gender disparities in science due to discrimination against women, family constraints that burden women's time, women's unwillingness to choose these careers and/or gender differences in ability. All committee members also rated women's and men's ability to be successful in their scientific field. The explicit belief that gender disparities are due to external barriers rather than internal abilities was assessed as the composite of ratings that women face discrimination and family constraints and do not lack ability, and are able to succeed in their field (further details about the procedure and measures are available in the Methods). Finally, 1 year later (when their participation in the study was likely to be less salient to them), all of the committee members met again to make their year 2 selection decisions with no explicit reminder about the study. Because the predicted effects should become more apparent when committees are less aware that their decision-making is being scrutinized, our analyses focused on changes in selections from the year the study was announced to the decisions made in the following year.

Selection decisions reflected committee-level consensus-based outcomes, thus all of the analyses below were performed on committee-level data. Committee members' implicit and explicit scores were averaged to index, respectively, the extremity of implicit associations and explicit beliefs for each committee (Supplementary Table 1). Because committee members were not assigned to committees based on their implicit associations or explicit beliefs, high intraclass correlation values on these variables were not expected. In fact, we make no assumption that the individual members of each committee would have inter-related science = male implicit associations or inter-related explicit beliefs about women in science. Rather, they come to the table with their distinct individual beliefs and implicit associations. We assumed that these explicit and implicit biases of individuals then shape the nature of the discussion and shared evaluation of the candidates that results in consensus-level decisions.

Results

The degree to which selection decisions disproportionately favour men or women was assessed primarily (both at year 1 and year 2) as an adverse impact (AI) ratio (see Methods), which takes into account the ratio of men and women in the applicant pool. AI values >1 reflect selections favouring women, whereas AI values <1 reflect decisions favouring men (1 = a selection ratio proportional to the gender ratio of the candidate pool). Because this measure is not symmetrical, it was log-transformed (see Methods). AI scores at years 1 and 2 were only modestly correlated (Pearson's $r = 0.35$), suggesting little stability in AI across the two years and motivating an interest in predicting changes in AI over time. We also used two alternative ways to compute committees' selection decisions: *d* score pass rates and gender asymmetry scores (see Methods). Table 1 presents descriptive statistics on selection data for years 1 and 2 separately, and the means of AI ratios, log-transformed AI ratios, *d* scores and gender asymmetry scores for all committees (see Methods).

A one-sample *t*-test against 0 (indicating no significant AI on the log-transformed AI scores) revealed no overall evidence of significant bias in selection decisions in either year 1 ($t(37) = 0.22$; $P = 0.826$; Cohen's $d = 0.036$; 95% confidence interval (CI): −0.07 to 0.09) or year 2 ($t(37) = -0.16$; $P = 0.871$; Cohen's $d = -0.027$; 95% CI: −0.08 to 0.07). However, we caution readers that given Simpson's paradox, the lack of overall bias when averaging across all committees does not imply that no selection bias is taking place within some committees. Committees that believe that external barriers constrain women's ability to succeed may show no relationship between implicit bias and selection outcomes, whereas those that minimize the existence of these barriers might show a stronger link between their implicit biases and selection decisions.

The aggregated IAT scores at the committee level revealed a significant implicit science = male association (one-sample Student's *t*-test against 0, $P < 0.001$), which was also present at the individual level (see Supplementary Information for individual-level data). The magnitude of this committee-level implicit gender–science association (IAT score: mean (M) = 0.36; s.d. = 0.15; 95% CI: 0.31 to 0.41) was similar to the average IAT score in a larger French sample[17] ($M = 0.42$; s.d. = 0.43; $n = 5{,}810$; 95% CI: 0.41 to 0.44), indicating that scientists themselves implicitly associate science more with men than with women. It is notable that without looking at the moderating effect of explicit attributions for women's under-representation, committee-level IAT scores were not significantly related to log-transformed AI ratios either in year 1 ($r = 0.16$; $P = 0.327$) or year 2 ($r = -0.06$; $P = 0.721$). This lack of relationship is not surprising given that implicit associations need not always relate to behaviour, but are more likely to predict behaviour when people feel justified to act on their biases. It is this theoretical perspective that informed the moderated analyses we conducted.

At the explicit level, committees on average expressed some belief that women face discrimination in science (committees' aggregated attributions to discrimination ranged from 2.56–4.50; $M = 3.64$; *d* score mean = 3.72; s.d. = 0.44; 1 = strongly disagree; 6 = strongly agree that women experience discrimination). However, nearly half of the committees tended to disagree that gender discrimination contributes to women's under-representation in STEM fields. Attributing disparities to gender discrimination correlated positively with the belief that family constraints burden women's research time ($r = 0.57$; $P < 0.001$) and negatively with the idea that gender differences in ability underlie women's under-representation in science ($r = -0.38$; $P = 0.017$; see Table 2). Given these intercorrelations, an approach to data reduction was useful to control for type I

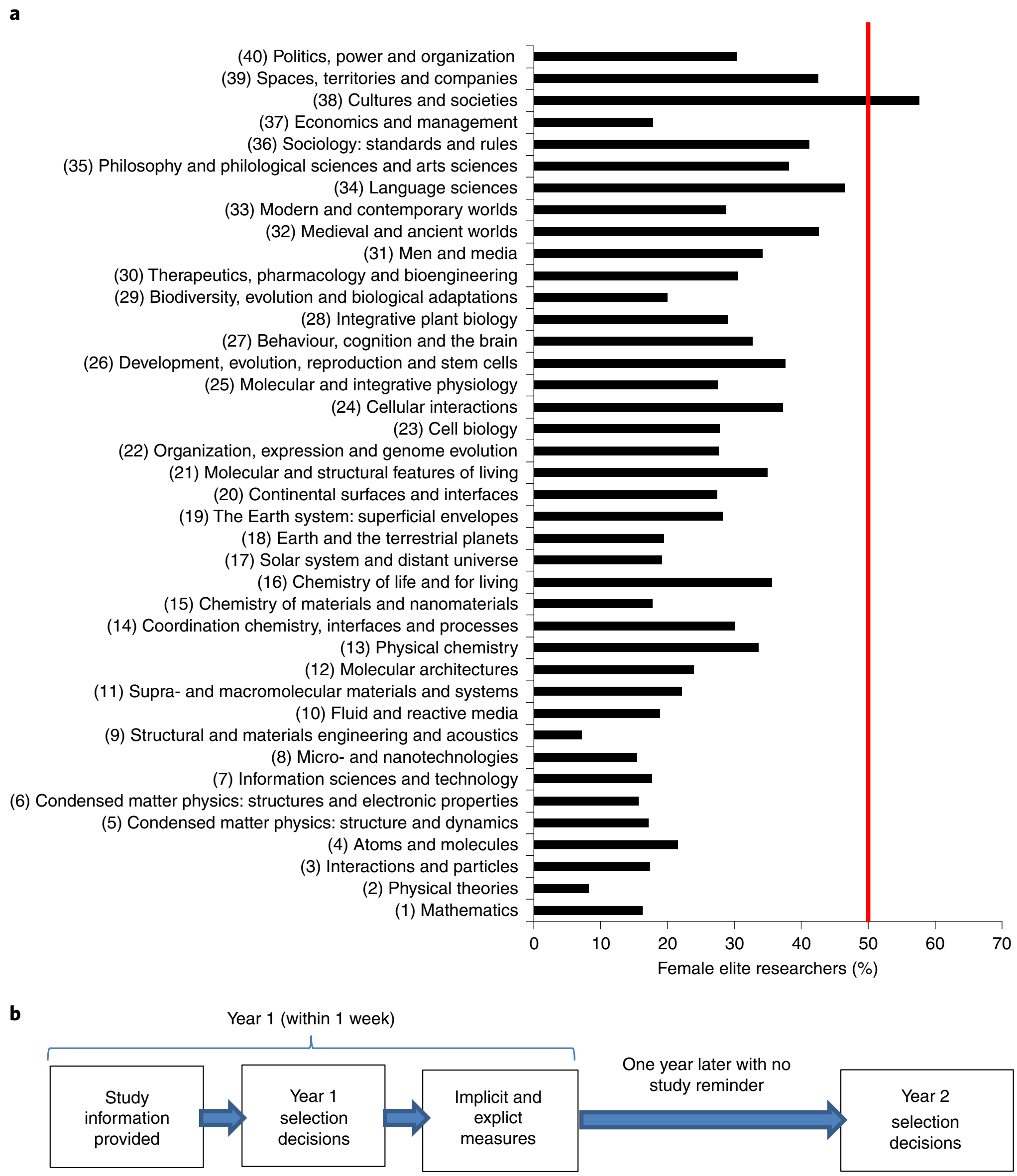

Fig. 1 | Gender asymmetry within each academic discipline in the year before the present study, and a timeline of the study. a, Percentage of female elite researchers in each discipline ($n = 4,759$ researchers across all disciplines). The red line indicates gender parity. The committees' structure is ordered by the governing body from the mathematics and physical science fields (numbers 1–20) to life and social sciences (numbers 21–40). Women are under-represented in all disciplines but one. **b**, The timeline was as follows: (1) immediately before the start of committee sessions, the committee members received preliminary information about the research; (2) committees started their work for selection decisions (at year 1), immediately followed by (3) completion of the IAT and questionnaire; (4) 1 year later, year 2 selection decisions were made but no explicit reminder of the study was made at that time.

errors in analysing several intercorrelated outcomes. A principal component factor analysis with varimax rotation conducted on the six explicit belief variables indicated that four of these indicators (discrimination, family constraints, women's low ability (reverse scored) and women's likelihood of success) loaded on a single factor that explained 40% of variance in responses (all factor loadings >0.50). Thus, higher values on the composite of these variables (Cronbach's $\alpha = 0.73$) reflected a stronger belief among committees that gender disparities are due more to external barriers (that is, women face discrimination and family constraints) rather than gender differences in ability (that is, women lack ability and are unlikely to succeed). Interestingly, there was no overall relationship between the committee-level measures of implicit associations and this composite belief that external barriers more than internal factors explain gender disparities in science (bivariate correlations; Table 2).

We then examined the interactive effect of both explicit and implicit gender biases on selection decisions. The necessity—in this real-world context—to inform the committees at year 1 about the aim of the research suggested that committee members would be more acutely aware of how their decisions were evaluated at year 1 than at year 2, and thus probably more cautious in their selection decisions at year 1 for social desirability purposes. As a result, we a priori focused on decisions made after a 1-year delay (year 2) as our core outcome variable. In addition, in the context of the organization's explicit interest in oversight and accountability, this study examined how group decisions could change over time (from year 1 to year 2) as a function of the beliefs and biases of constituent group members. Thus, the hypothesis tested by our model is whether committees with stronger implicit associations and who do not endorse external barriers as a problem exhibit the largest decrease in

Table 1 | Selection data for years 1 and 2, reporting the size of the candidate pool, percentage of women in the pool, number of chosen men and women, and the three ways of operationalizing selection decisions (AI ratios, *d* scores and gender asymmetry scores)

Committee	Year 1							Year 2						
	Pool	% female	Chosen men	Chosen women	AI ratio (log[AI ratio])	*d* score	Gender asymmetry (men – women)	Pool	% female	Chosen men	Chosen women	AI ratio (log[AI ratio])	*d* score	Gender asymmetry (men – women)
1	55	16	9	1	0.57 (−0.25)	−13.80	8	49	14	5	2	2.40 (0.38)	19.53	3
2	39	10	8	1	1.09 (0.04)	1.39	7	33	21	6	3	1.86 (0.27)	12.80	3
3	58	28	11	7	1.67 (0.22)	20.87	4	62	27	12	3	0.66 (−0.18)	−14.80	9
4	51	24	9	3	1.08 (0.03)	2.20	6	48	23	10	2	0.67 (−0.17)	−10.42	8
5	48	23	10	2	0.67 (−0.17)	−10.42	8	45	20	9	1	0.44 (−0.35)	−18.37	8
6	28	14	7	3	2.57 (0.41)	22.82	4	29	21	9	1	0.43 (−0.37)	−14.51	8
7	65	26	12	7	1.65 (0.22)	21.68	5	65	23	13	4	1.03 (0.01)	0.94	9
8	50	20	10	3	1.20 (0.08)	5.04	7	48	19	9	2	0.96 (−0.02)	−0.91	7
9	24	21	8	1	0.48 (−0.32)	−10.91	7	28	14	9	0	0 (-)	—	9
10	45	27	13	3	0.63 (−0.20)	−13.54	10	41	29	11	4	0.88 (−0.06)	−3.75	7
11	33	33	5	4	1.60 (0.20)	10.52	1	30	30	6	3	1.17 (0.07)	3.07	3
12	23	26	6	2	0.94 (−0.02)	−0.88	4	22	9	7	1	1.43 (0.15)	3.98	6
13	43	44	2	7	4.42 (0.65)	45.66	−5	42	31	6	4	1.49 (0.17)	10.47	2
14	42	43	7	4	0.76 (−0.12)	−7.64	3	37	49	5	4	0.84 (−0.07)	−4.14	1
15	31	32	8	3	0.79 (−0.10)	−5.27	5	34	32	6	3	1.05 (0.02)	0.96	3
16	54	50	6	6	1.00 (0)	0	0	46	52	6	6	0.92 (−0.04)	−2.71	0
17	51	24	10	2	0.65 (−0.19)	−11.93	8	42	24	8	2	0.80 (−0.10)	−5.16	6
18	35	26	8	2	0.72 (−0.14)	−6.81	6	26	27	7	2	0.78 (−0.11)	−4.38	5
19	33	33	7	5	1.43 (0.15)	9.15	2	26	31	6	4	1.50 (0.18)	8.49	2
20	20	30	3	3	2.33 (0.37)	12.76	0	22	23	5	3	2.04 (0.31)	12.63	2
21	40	40	8	6	1.13 (0.05)	3.58	2	43	49	7	5	0.75 (−0.13)	−8.68	2
22	38	55	11	6	0.44 (−0.36)	−31.63	5	35	57	8	9	0.84 (−0.07)	−5.82	−1
23	45	58	6	8	0.97 (−0.01)	−0.84	−2	33	64	3	12	2.29 (0.36)	23.81	−9
24	41	41	12	6	0.71 (−0.15)	−12.43	6	37	49	11	5	0.48 (−0.32)	−25.19	6
25	32	22	9	5	1.98 (0.30)	21.61	4	26	42	6	6	1.36 (0.13)	7.66	0
26	22	50	8	4	0.50 (−0.30)	−18.67	4	25	40	8	6	1.13 (0.05)	3.34	2
27	31	29	8	5	1.53 (0.18)	11.30	3	34	44	6	5	1.06 (0.02)	1.35	1
28	14	50	6	3	0.50 (−0.30)	−17.32	3	13	38	6	2	0.53 (−0.27)	−10.43	4
29	22	41	7	5	1.03 (0.01)	0.75	2	27	33	11	6	1.09 (0.04)	3.08	5
30	44	36	8	6	1.31 (0.12)	8.64	2	33	42	7	4	0.78 (−0.11)	−6.19	3
31	26	46	6	7	1.36 (0.13)	8.22	−1	21	52	3	6	1.82 (0.26)	11.28	−3
32	28	54	4	6	1.30 (0.11)	5.70	−2	20	50	4	6	1.50 (0.18)	8.33	−2
33	17	59	5	3	0.42 (−0.38)	−16.74	2	17	65	4	2	0.27 (−0.56)	−21.66	2
34	12	42	2	3	2.10 (0.32)	8.40	−1	13	46	2	3	1.75 (0.24)	6.15	−1
35	24	29	7	4	1.39 (0.14)	7.15	3	18	33	7	3	0.86 (−0.07)	−2.87	4
36	28	54	8	8	0.87 (−0.06)	−4.72	0	21	48	7	6	0.94 (−0.03)	−1.62	1
37	20	50	5	3	0.60 (−0.22)	−8.73	2	24	38	5	4	1.33 (0.12)	5.51	1
38	27	74	4	7	0.61 (−0.21)	−11.10	−3	21	81	3	6	0.47 (−0.33)	−15.82	−3
39	10	0	8	0	—	—	8	12	17	4	1	1.25 (0.10)	1.83	3
40	16	38	4	4	1.67 (0.22)	8.94	0	12	25	6	2	1.00 (0)	0	4
Mean	34.12	35.45	7.38	4.20	1.20 (0.01)	0.85	3.18	32.50	35.80	6.83	3.82	1.07 (−0.01)	−0.83	3.00
s.d.	13.72	15.33	2.64	2.10	0.76 (0.24)	14.55	3.46	12.95	15.90	2.59	2.34	0.54 (0.21)	10.74	3.81
n					39 (39)	39	40					40 (39)	39	40

selecting women for these positions (that is, a decrease in AI ratios) at year 2 relative to year 1. In contrast, committees with a stronger implicit association but who endorse external barriers as a problem might be motivated to similarly suppress their implicit biases in both years (no change) or counteract biases in decisions made in year 1 by selecting more women in year 2 (that is, an increase in AI ratios).

Table 2 | Pearson's bivariate correlations (exact *P* values) between committee-level variables in the study

Variable	IAT	External barriers (composite)	Discrimination	Family constraints	Personal choice	Ability	Women's success	Men's success	Year 1 log[AI ratio]
IAT	–								
External barriers (composite)	0.21 (0.195)								
Discrimination	0.05 (0.753)	0.72 (0.001)							
Family constraints	0.08 (0.641)	0.81 (0.001)	0.57 (0.001)						
Personal choice	0.26 (0.106)	−0.25 (0.122)	−0.28 (0.083)	−0.24 (0.142)					
Ability	−0.09 (0.578)	−0.73 (0.001)	−0.38 (0.017)	−0.49 (0.002)	0.16 (0.336)				
Women's success	0.34 (0.032)	0.69 (0.001)	0.15 (0.376)	0.30 (0.064)	−0.06 (0.695)	−0.51 (0.001)			
Men's success	−0.21 (0.202)	0.24 (0.139)	0.20 (0.230)	0.36 (0.025)	0.04 (0.828)	−0.17 (0.303)	0 (0.986)		
Year 1 log[AI ratio]	0.16 (0.327)	−0.21 (0.203)	−0.43 (0.007)	−0.03 (0.873)	−0.07 (0.698)	0.27 (0.106)	0.05 (0.749)	−0.21 (0.216)	
Year 2 log[AI ratio]	−0.06 (0.721)	−0.14 (0.415)	−0.19 (0.248)	−0.11 (0.521)	−0.06 (0.731)	0.17 (0.321)	0.02 (0.912)	−0.10 (0.534)	0.35 (0.032)

$n = 38$–40. The composite score of external barriers is the committee-level average of attribution to discrimination, family constraints, gender differences in ability (reverse scored before analyses) and perceived likelihood of women's success (Cronbach's $\alpha = 0.73$).

Statistically, this means that year 1 selection decisions are thought to be acting as a suppressor variable that is confounded slightly with both our predictor (committees who selected more women in year 1 are more likely to doubt that barriers hold women back; $r = -0.21$) and our outcome (AI scores in both years are somewhat positively correlated; $r = 0.35$). Thus, the model tested assumes a prediction chain where: explicit beliefs ← year 1 AI → year 2 AI. We tested a regressor variable model (that is, the residual gain score model) that uses year 1 AI as a covariate, as not controlling for these suppression effects would lead to bias in the estimated explicit beliefs by IAT interaction effect on year 2 AI. In Table 3, we provide a summary of the key results using other analytical strategies for measuring and analysing selection decisions.

A multiple regression analysis[25] (Methods) tested the key expected interactive effect between committee-level implicit science = male associations and explicit beliefs about external barriers on committees' decisions at year 2, while controlling for selection decisions from the previous year (log-transformed AIs). The results revealed that this interaction effect was significant ($\beta = 0.38$; 95% CI: 0.04 to 0.72; $t(32) = 2.30$; $P = 0.028$). Table 3 (part a1) shows that, as expected, committees with a stronger science = male association exhibited the largest decrease in selecting women (a lower log-transformed AI ratio) if those committees also had weaker beliefs that external barriers hold women back (16th percentile; $\beta = -0.61$; 95% CI −1.13 to −0.10; $t(32) = -2.42$; $P = 0.021$). Implicit gender bias was unrelated to selection decisions in those committees whose members believed that gender disparities in science can be due to external barriers (84th percentile; $\beta = 0.13$; 95% CI: −0.31 to 0.56; $t(32) = 0.60$; $P = 0.555$). The predicted interaction was also significant: (1) with raw AI scores (Table 3, part a2); (2) when using alternative measures of committees' selection decisions (*d* score pass rate and gender asymmetry scores; Table 3, parts a3 and a4); (3) with a change score analysis using the difference score between year 2 AI and year 1 AI (Table 3, parts b1–4); and (4) while controlling for other covariates, such as the percentage of women in the evaluation committee and designation of the committee as a mathematics/physical science versus social/life science (Supplementary Table 2).

Effect sizes are similar in direction but weaker in magnitude in less sensitive analyses that do not control for selection decisions at year 1 (Table 3, part c). The hypothesized interaction effect was significant when analysing year 2 raw AI scores (Table 3, part c2; $\beta = 0.29$; 95% CI: 0.04 to 0.53; $t(32) = 2.36$; $P = 0.024$), but was not significant for the other selection variables (Table 3, parts c1, c3 and c4). Finally, for transparency in reporting, there were no significant main effects or interaction of implicit associations and explicit beliefs on year 1 selection decisions that were made just before the measurement of our predictor variables (Table 3, parts d1–4). Descriptively, there was a numerical trend for committees with stronger implicit biases, paired with a lower belief that barriers are a problem, to initially favour women in their selection decisions made at year 1 when the study of gender bias was first announced (Supplementary Fig. 1a). These committees showed the opposite numerical trend at year 2 when not under scrutiny (Supplementary Fig. 1b). Neither of these findings was statistically significant; however, they further supported our motivation to analyse change in selection decisions from year 1 to year 2. Committees who endorsed external barriers as a problem made decisions closer to parity at both years, whatever their implicit bias (Supplementary Fig. 1a–c).

Discussion

Many factors contribute to women's under-representation in scientific fields[26–29]. The present research highlights that decision-makers' beliefs about these disparities might contribute to the barriers women face. The findings from this real-world study suggest that committees might be more likely to act on their implicit gender biases when, at an explicit level, they do not strongly believe that systemic biases are a problem that need to be addressed. Under these conditions, committees are less likely to select accomplished women for elite research positions. These effects seemed to become evident 1 year after the study was announced when committee members probably felt less externally scrutinized for biases in their decision-making. By highlighting when implicit associations do and do not predict selection decisions, these results also help to reconcile inconsistencies in past research. Depending on the degree to which they believe that external barriers hold women back, committees may or may not break the habit of implicit gender bias. These findings are noteworthy and unique because they come from data collected in the field as part of a nationwide

Table 3 | Comparison of the results for the focal implicit × explicit belief interaction tested on AI ratios (with and without log-transformation), *d* score pass rates and gender asymmetry scores

	Barrier × IAT interaction β (s.e.); P; 95% CI	Simple slope at low explicit belief β (s.e.); P; 95% CI	Simple slope at high explicit belief β (s.e.); P; 95% CI
(a) Covariate analyses on year 2 selection decisions with year 1 selection decisions as a covariate			
(1) Log-transformed year 2 AI with year 1 AI as a covariate	$\beta=0.38$ (0.17); $t=2.30$; $P=0.028$; 95% CI: 0.04 to 0.72	$\beta=-0.61$ (0.25); $t=-2.42$; $P=0.022$; 95% CI: −1.13 to −0.10	$\beta=0.13$ (0.21); $t=0.60$; $P=0.555$; 95% CI: −0.31 to 0.56
(2) Raw year 2 AI with year 1 AI as a covariate	$\beta=0.49$ (0.16); $t=3.12$; $P=0.004$; 95% CI: 0.17 to 0.81	$\beta=-0.83$ (0.23); $t=-3.61$; $P=0.001$; 95% CI: −1.30 to −0.36	$\beta=0.10$ (0.20); $t=0.48$; $P=0.634$; 95% CI: −0.31 to 0.51
(3) Year 2 *d* score pass through with year 1 AI as a covariate	$\beta=0.38$ (0.18); $t=2.13$; $P=0.041$; 95% CI: 0.02 to 0.74	$\beta=-0.59$ (0.27); $t=-2.16$; $P=0.038$; 95% CI: −1.15 to −0.03	$\beta=0.14$ (0.22); $t=0.63$; $P=0.533$; 95% CI: −0.31 to 0.60
(4) Year 2 asymmetry scores with year 1 AI as a covariate	$\beta=-0.22$ (0.09); $t=2.49$; $P=0.018$; 95% CI: −0.40 to −0.04	$\beta=0.40$ (0.16); $t=2.53$; $P=0.016$; 95% CI: 0.08 to 0.73	$\beta=-0.02$ (0.12); $t=-0.20$; $P=0.845$; 95% CI: −0.28 to 0.23
(b) Change score analyses			
(1) Log-transformed year 2 AI change scores	$\beta=0.40$ (0.17); $t=2.45$; $P=0.020$; 95% CI: 0.07 to 0.74	$\beta=-0.74$ (0.24); $t=-3.04$; $P=0.005$; 95% CI: −1.24 to −0.25	$\beta=0.04$ (0.21); $t=0.18$; $P=0.860$; 95% CI: −0.40 to 0.47
(2) Raw year 2 AI change scores	$\beta=0.45$ (0.15); $t=2.98$; $P=0.005$; 95% CI: 0.14 to 0.76	$\beta=-0.84$ (0.22); $t=-3.88$; $P<0.001$; 95% CI: −1.29 to −0.40	$\beta=0.02$ (0.20); $t=0.08$; $P=0.936$; 95% CI: −0.39 to 0.42
(3) Year 2 *d* score change scores	$\beta=0.41$ (0.16); $t=2.53$; $P=0.016$; 95% CI: 0.08 to 0.74	$\beta=-0.77$ (0.24); $t=-3.18$; $P=0.003$; 95% CI: −1.26 to −0.28	$\beta=0.03$ (0.21); $t=0.14$; $P=0.885$; 95% CI: −0.40 to 0.46
(4) Year 2 asymmetry change scores	$\beta=-0.35$ (0.11); $t=-3.20$; $P=0.003$; 95% CI: −0.57 to −0.13	$\beta=0.62$ (0.20); $t=3.13$; $P=0.004$; 95% CI: 0.22 to 1.03	$\beta=-0.06$ (0.16); $t=-0.34$; $P=0.732$; 95% CI: −0.38 to 0.27
(c) Analyses on year 2 selection decisions without year 1 selection decisions as a covariate			
(1) Log-transformed year 2 AI without a covariate	$\beta=0.25$ (0.13); $t=2.00$; $P=0.054$; 95% CI: −0.01 to 0.51	$\beta=-0.36$ (0.24); $t=-1.55$; $P=0.131$; 95% CI: −0.85 to 0.11	$\beta=0.14$ (0.19); $t=0.74$; $P=0.463$; 95% CI: −0.24 to 0.51
(2) Raw year 2 AI without a covariate	$\beta=0.29$ (0.12); $t=2.36$; $P=0.024$; 95% CI: 0.04 to 0.53	$\beta=-0.54$ (0.22); $t=-2.42$; $P=0.021$; 95% CI: −0.99 to −0.08	$\beta=0.02$ (0.18); $t=0.12$; $P=0.906$; 95% CI: −0.34 to 0.38
(3) Year 2 *d* score pass through without a covariate	$\beta=0.21$ (0.13); $t=1.60$; $P=0.120$; 95% CI: −0.05 to 0.47	$\beta=-0.31$ (0.24); $t=-1.25$; $P=0.219$; 95% CI: −0.80 to 0.19	$\beta=0.11$ (0.19); $t=0.57$; $P=0.571$; 95% CI: −0.28 to 0.49
(4) Year 2 asymmetry scores without a covariate	$\beta=-0.04$ (0.12); $t=-0.37$; $P=0.717$; 95% CI: −0.29 to 0.20	$\beta=0.12$ (0.22); $t=0.52$; $P=0.604$; 95% CI: −0.34 to 0.57	$\beta=0.03$ (0.18); $t=0.17$; $P=0.869$; 95% CI: −0.34 to 0.39
(d) Analyses on year 1 selection decisions			
(1) Log-transformed year 1 AI	$\beta=-0.10$ (0.18); $t=-0.53$; $P=0.596$; 95% CI: −0.46 to 0.27	$\beta=0.29$ (0.25); $t=1.14$; $P=0.261$; 95% CI: −0.23 to 0.81	$\beta=0.11$ (0.23); $t=0.64$; $P=0.463$; 95% CI: −0.36 to 0.58
(2) Raw year 1 AI	$\beta=-0.19$ (0.17); $t=-1.11$; $P=0.275$; 95% CI: −0.54 to 0.16	$\beta=0.44$ (0.25); $t=1.77$, $P=0.085$; 95% CI: −0.06 to 0.94	$\beta=0.07$ (0.23); $t=0.12$; $P=0.331$; 95% CI: −0.38 to 0.53
(3) Year 1 *d* score pass through	$\beta=-0.19$ (0.17); $t=-1.07$; $P=0.291$; 95% CI: −0. 54 to 0.17	$\beta=0.44$ (0.25); $t=1.79$; $P=0.082$; 95% CI: −0.06 to 0.94	$\beta=0.09$ (0.22); $t=0.40$; $P=0.691$; 95% CI: −0.37 to 0.55
(4) Year 1 asymmetry scores	$\beta=0.24$ (0.13); $t=1.93$; $P=0.062$; 95% CI: −0.01 to 0.50	$\beta=-0.40$ (0.23); $t=-1.73$; $P=0.093$; 95% CI: −0.87 to 0.07	$\beta=0.08$ (0.19); $t=0.41$; $P=0.686$; 95% CI: −0.30 to 0.45

All variables are standardized in the models. The composite score of external barriers is the committee-level average of attribution to discrimination, family constraints, gender differences in ability (reverse scored before analyses) and perceived likelihood of women's success (Cronbach's $\alpha=0.73$). Asymmetry is coded so that higher numbers mean favouring men over women, whereas AI and *d* score pass through rates are interpreted as lower numbers imply favouring men over women. Simple slopes are estimated at the 16th (low) and 84th (high) percentile of the distribution for each moderator.

competition for elite research positions with real consequences for female and male scientists.

At the same time, the findings from this single correlational study should be interpreted with caution given that a study of this kind does not afford causal inferences (it would not be feasible to carry out a randomized controlled trial during the natural course of a nationwide competition of this kind). Given the correlational nature of the research, it is of course possible that the relationships observed here are better explained by another variable that was not assessed. In addition, our sample size was limited by the fact that there are only 40 committees (39 with complete data); thus, the study is underpowered to detect the key interaction effect. However, it is also worth noting that the high stakes of this real-world evaluation could enhance careful responding and increase measurement precision, which should maximize the ability to detect true effects in a study of this type. As a check against type 1 error, we note that the key effect is generally significant across different analytical approaches, but not terribly strong. A replication with other evaluation committees in real-world competitions is necessary to ascertain the strength and generalizability of these findings. Thus, the present study should be viewed as an initial effort to document the interactive effect of both explicit beliefs and implicit gender biases on hiring and promotion decisions in the real world of science. Although inherent in a field study, it is essential that limitations are addressed in future research.

It was still common in the last century to explicitly contest prestigious scientific positions awarded to women, as was the case with the appointment of Marie Curie to the French Academy of Sciences. Given the present evidence that gender bias can still exist today in academic science—at least at the implicit level—we highlight the need for efforts to educate committees and governing bodies about the existence and consequences of these biases. In this research, committees who acknowledged that biases can exist were less likely to show any link between their implicit biases and selection outcomes. Recognizing the role that such biases can play might enable committees to set them aside at the time of final decisions, thereby facilitating gender equity and diversity[30]. As such, the present findings support 'habit-breaking interventions' that involve[14]: (1) making committee members aware of implicit biases; (2) making them able to understand the consequences of these biases; and (3) providing them with effective strategies to reduce the impact of implicit biases. Future research could specify whether this three-step intervention can maximize accurate decision-making among those committees that hold implicit gender biases, while simultaneously doubting that external barriers contribute to women's under-representation in STEM fields. Any evidence in this direction would help persuade future evaluation committees to be mindful of their biases when making promotion decisions in the real world. As suggested by the present research, even committees whose members hold strong gender biases might be prevented from acting on them when they feel more accountable for making unbiased decisions (here, at year 1) but might also exhibit reactance when no longer scrutinized (here, at year 2). The efficiency of educating committees about gender biases may therefore be maximized when combined with strong accountability measures.

Methods

Participants. All of the members of the National Committee for Scientific Research, which plays a key role in French science, were encouraged by the governing body to participate in a study on women's under-representation in science. This national committee is a collective body comprising a general Scientific Board, ten Institute Scientific Boards and (at the time of the present study) 40 specialized committees that cover the entire scientific spectrum. As indicated in the caption to Fig. 1, the committees' structure is ordered by the governing body from mathematics and physical sciences (numbers 1–20) to life/social sciences (numbers 21–40). Members of these specialized committees (about 20 researchers per committee) meet three times a year, typically for 3–4 d. During these different sessions, they adjudicate the selection of junior and senior accomplished researchers for promotion to more advanced research positions, monitor previously recruited researchers and laboratory activities, and identify (and later carry out, either alone or with partners) all research that advances science or contributes to the country's economic, social and cultural progress. Overall, these committees manage around 20,000 scientific files each year, and their members are renewed every 4 years.

The present study took place half-way through the committees' mandate, with 426 members from 39 of the 40 specialized committees volunteering to participate in the study. Committees members ranged in age from 35–64 years (note that age is not included in the database to preserve evaluators' anonymity). Due to IAT errors (see 'Measures' below), data from 12 members were excluded from the dataset, resulting in a total of 414 members (254 men, 154 women and six of unspecified gender) representing 50% of the whole population. This participation rate is relatively high considering that participants came from the real world and their decisions would have real consequences for the candidates' actual careers. In contrast with people participating in mock-hiring scenarios, real committee members may be reluctant to report on their attitudes and beliefs about women's under-representation in science. Yet, our 50% rate of participation is greater than the 30–35% response rate in previous high-profile experimental investigations of gender bias in selection decisions using academic samples[4,7]. The study received ethical review and was approved by the Centre National de la Recherche Scientifique governing body at the time of the current study, as well as the Mission for Women's Integration and the National Committee of Scientific Research (CoNRS General Secretary). Informed consent was obtained from all participants and no compensation was provided for participating in the study.

Procedure. The study was conducted over 2 years. During year 1, after the competition for research positions was completed, committee members met again for other activities and were reminded by the governing body of the possibility of participating in a study (preserving anonymity) on women's under-representation in science. They were told that the study was aimed at examining whether gender biases contribute to women's under-representation in science, but it was not made explicit that their survey and IAT data would be used to predict committee-level selection decisions. They were invited to complete the gender–science IAT along with measures of their perceptions of the origin of current gender disparities in STEM fields due to discrimination, ability, family constraints or personal choice, and their perceptions of men's and women's likelihood of achieving success in their own scientific field. The use of both implicit and explicit measures was motivated by results suggesting that the IAT does not necessarily have independent predictive validity but rather should be moderated by one's explicit beliefs and values[14,31]. The gender–science IAT and questionnaire ratings were counterbalanced across participants and made accessible in French language on an adapted version of Greenwald, Banaji and Nosek's Project Implicit web platform.

Each participant had free access to a computer throughout the day in a room adjoining the National Committee rooms, where they worked alone (15 min in total) during their breaks and other free times. Two members of the research team were present to address technical difficulties, but had minimal to no interaction with participants. The participation of the committee members was limited to performing the gender–science IAT and filling out the questionnaires at year 1. At year 2, there was no reminder of the study and no measures were directly taken among participants. Both year 1 and year 2 selection decisions were given to our research team directly by the governing body. Thus, participants had virtually no direct contact with the research team, as explicit data were collected online and selection decision data were provided to the research team by the governing body.

Measures. *AI ratio.* Each committee's decision outcomes for years 1 and 2 (that is, the number of men and women selected and in the candidate pools) were used to compute an AI ratio that compares the pass rates of men and women while taking into account the number of men and women in the applicant pool. The following formulas were used to compute an AI score for each committee in each year:

$$P_{\mathrm{f}} = n_{\mathrm{f}}/N_{\mathrm{f}}$$

$$P_{\mathrm{m}} = n_{\mathrm{m}}/N_{\mathrm{m}}$$

$$\mathrm{AI} = P_{\mathrm{f}}/P_{\mathrm{m}}$$

where P_{f} and P_{m}, respectively, are the probabilities of females and males being selected, n_{f} and n_{m}, respectively, are the numbers of females and males selected, and N_{f} and N_{m}, respectively, are the numbers of females and males in the applicant pool. AI values >1 reflect decisions that disproportionately favour women, whereas AI values <1 reflect decisions that disproportionately favour men. An AI of 1 reflects a gender ratio in decisions that is exactly proportional to the ratio in the candidate pool. This measure is not symmetrical: for all cases where the probability of women to be selected is smaller than the probability for men to be selected (that is, $P_{\mathrm{f}} < P_{\mathrm{m}}$), AI will lie between 0 and 1. In contrast, for all cases where the probability of women to be selected is larger than that of men (that is, $P_{\mathrm{f}} > P_{\mathrm{m}}$), AI will lie between 1 and infinity. This asymmetry was solved by taking the logarithm of the ratio of probabilities since $\log[P_{\mathrm{f}}/P_{\mathrm{m}}] = -\log[P_{\mathrm{m}}/P_{\mathrm{f}}]$. We report analyses on the log of AI. We also present the results of our analyses without the log-transformation, given that AI is typically analysed without transformation[32] (see Table 3, part a2).

Alternative measures of committee's selection decisions. We used two other ways to compute gender bias in selection decisions: *d* score pass rates and gender asymmetry scores. The *d* score is an effect size for each committee in each year of data, and was calculated as follows:

$$d = \frac{(P_{\mathrm{f}} - P_{\mathrm{m}})}{\sqrt{\left(\frac{P_{\mathrm{f}}(1-P_{\mathrm{f}})}{n_{\mathrm{f}}}\right)\left(\frac{P_{\mathrm{m}}(1-P_{\mathrm{m}})}{n_{\mathrm{m}}}\right)}}$$

A d score of 0 represents no gender asymmetry, and negative values represent bias against women. The *d* score pass rate was highly related to the log-transformed AI ratio, both for year 1 ($r = 0.96$; $P < 0.001$) and year 2 selections ($r = 0.96$; $P < 0.001$).

Gender asymmetry scores were computed as the difference between the number of males and the number of females selected by a committee for a particular year (a positive difference score indicating women's under-representation). Gender asymmetry was significantly correlated—but not redundant—with the log-transformed AI ratios from year 1 ($r = -0.45$; $P = 0.004$) and year 2 ($r = -0.43$; $P = 0.007$). Although these raw asymmetry scores do not take into account the gender ratio in the applicant pool, they might be meaningful given that committees who are concerned with mitigating gender disparities might explicitly focus on the gender ratio in selections.

IAT. We used the French version of the gender–science IAT[15,17]. This test measures the association strength between the concepts 'male' and 'female' and the attributes 'science' and 'liberal arts.' Its structure is a within-subject experiment involving two conditions in which the pairings of these four categories are varied. Words representing the four categories are presented one at a time in the centre of the computer screen, and participants categorize each by pressing one of two keys. In one condition, participants categorize male and science words with one key, and female and liberal arts words with the other key. In the other condition, participants categorize female and science words with one key, and male and liberal arts words with the other key. The order of these conditions is randomized across participants. The difference in average categorization latency between the two conditions is an indicator of association strength between the gender and academic categories. Here, the 'stereotype-congruent' condition is when male and science words share a response key and female and liberal arts words share the other. Faster categorization in this condition compared with the 'stereotype-incongruent' condition (when male and liberal arts words share a response key and female and science words share the other) indicates stronger associations of male with science and female with liberal arts compared with the reverse. Following Greenwald et al.[16], effect-size *d* scores were computed for each participant by dividing the difference in mean response latency between the two IAT conditions by the participant's latency standard deviation inclusive of the two conditions.

The IAT procedure followed the standard method described by Nosek et al.[33], and data were analysed using their improved scoring algorithm with the following features: responses faster than 400 ms were removed; responses slower than 10,000 ms were removed; and errors were replaced with the mean of the correct responses in that response block plus a 600-ms penalty. In addition to the data-cleaning procedures, IAT scores were disqualified for any of the following criteria suggestive of careless participation: (1) going too fast (<300 ms) on more than 10% of the total test trials; (2) 25% of responses too fast in any one of the critical blocks; (3) 35% too fast in any one of the practice blocks; (4) making more than 30% erroneous responses across the critical blocks; (5) 40% errors in any one of the critical blocks; (6) 40% errors across all of the practice blocks; or (7) 50% errors in any one of the practice blocks. These standards resulted in a disqualification rate of 2.8% of respondents.

Self-reports (questionnaire). The questionnaire assessed participants' perceptions of the origin of women's under-representation in STEM fields, and their perceptions of men's and women's success in different scientific disciplines. Participants were first reminded that women are under-represented in certain scientific fields, such as chemistry, physical sciences, mathematics, engineering and astronomy. Using a six-point scale (1 = strongly disagree; 2 = disagree; 3 = somewhat disagree; 4 = somewhat agree; 5 = agree; 6 = strongly agree), participants were then asked to indicate the extent to which they personally agreed or disagreed that each of the following factors contribute to women's under-representation in STEM fields: discrimination; gender differences in ability in these fields; family constraints; and personal choice. The eight items used to assess these variables were: (1) discrimination: "On average, with an equivalent scientific record, men are nevertheless advantaged over women in recruitment and promotion processes"; "Whatever their scientific abilities, women are often discriminated against"; and "On average, women are encouraged less than are men to take on management responsibilities (teams, laboratories, major programs, etc.)"; (2) gender differences in ability: "On average, men and women differ in their ability to exercise leadership responsibilities (teams, laboratories, major programs, etc.)'; 'On average, men and women do not have the same scientific abilities"; (3) family constraints: "On average, women are forced to invest more than men in their family/private lives, possibly to the detriment of their working lives"; and (4) personal choice: "On average, men and women differ in their willingness to assume management responsibilities (teams, laboratories, major programs, etc.)"; and "On average, women deliberately choose to invest more than men in their family/private life, possibly to the detriment of their working lives". For all items, lower (higher) scores indicate disagreement (agreement) with the proposed factors as an explanation for women's under-representation in STEM fields.

A total of 18 additional items measured participants' expectations of men's and women's success in mathematics, physics, chemistry, engineering, information sciences, earth sciences and astronomy, biological sciences, ecology and environment, humanity and social sciences. For each field, participants indicated the likelihood of success for men and women separately using a seven-point scale from 1 (very unlikely) to 7 (very likely). Analyses focused on participants' ratings of men and women only for their own discipline.

Estimates of implicit bias and explicit beliefs at the committee level. Supplementary Table 1 contains the committee-level measures of implicit science = male associations and other explicit beliefs created by averaging the responses of individual committee members for each variable. The committee-level aggregate of attributions to ability was reverse-scored so that higher numbers reflect a rejection of the idea that gender disparities are due to gender differences in ability. IAT scores and other explicit measures were considered additive or compositional properties of each committee[34,35]. In contrast with consensus variables, there was no expectation that individual members of each committee would have inter-related implicit associations about gender and science or inter-related explicit beliefs about discrimination. These measures should not be interpreted as the beliefs of the group as an entity, but rather as an estimate of the extremity of implicit associations and explicit beliefs that could be inputs to the group discussion of candidates in the process of reaching group consensus. A similar approach has been used in studies of how the average personality or intelligence scores of group members predict team performance outcomes[36–39].

Statistical analyses. Since selection decisions were made at the committee level, most statistical analyses were performed on committee-level data. Some analyses were nonetheless performed at the individual data level, for descriptive purposes (see Supplementary information for more detail). We used an α level of 0.05 (two tailed) for all statistical tests.

Descriptive statistics and gender differences on individual-level data. We used a one-sample t-test with a comparison against 0 (two tailed) to estimate the magnitude of the implicit science = male bias across individuals, and independent sample t-tests (two tailed) to test gender differences on each implicit and explicit measure.

Descriptive statistics on committee-level data. We also used one-sample t-tests with a comparison against 0 (two tailed) to estimate the magnitude of the implicit science = male bias (IAT score) across committees, and Pearson's bivariate correlations were used to examine overall relationships between implicit bias, explicit ratings and selection decisions.

Analytical strategy for the test of key hypotheses. We hypothesized that a strong science = male bias would predict a greater gender asymmetry in selections for those committees that do not strongly believe that external barriers constrain women's advancement in STEM. In contrast, committees whose members believe that gender disparities can be due to external barriers were expected to show a significantly weaker relationship between implicit bias and gender asymmetrical selections, consistent with the idea that they are more motivated to suppress or even counteract the role that biases may play in their decisions. Given that committees were likely to feel greater scrutiny for their selections at year 1 compared with year 2, analyses focused on analysing committees' change in selection decisions over the year of the study. To test our hypothesis, we carried out a moderated regression analysis using the Process macro designed by Hayes[25]. This analytical tool relies on ordinary least squares for estimating interactions in multiple regression along with simple slopes and regions of significance for probing interactions. The data met key assumptions of moderated linear regression. There was no evidence for curvilinear effects of either the predictor or the moderator on the outcome variable. All variables were standardized to achieve equal variance and avoid multicollinearity between predictors and the interaction term in the model. Predictors were never correlated with one another above $r = 0.50$, and multicollinearity statistics were all within an acceptable range.

In the analyses, year 2 AI scores were regressed on committee-level IAT bias, the committee-level attributions to external barriers (composite variable) and the interaction of these two predictors, while controlling for year 1 AI scores. All variables in the model were standardized.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Most of the data supporting the findings of this study at the committee level are available within the paper (and its Supplementary Information files). The entire dataset that supports the findings of this study is available from the Open Science Framework repository (https://osf.io/umf62/). Individual-level data are available on request from the authors.

Code availability

The code used to perform the primary analyses of the study is available from the corresponding authors upon request.

Received: 21 September 2017; Accepted: 10 July 2019;
Published online: 26 August 2019

References

1. Gibney, E. Women under-represented in world's science academies. *Nature News* http://www.nature.com/news/women-under-represented-in-world-s-science-academies-1.19465 (29 February 2016).
2. *Gender in research and innovation. Statistics and Indicators* https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-leaflet-web.pdf (She Figures, European Union, 2015).
3. *Science and Engineering Indicators 2016* https://www.nsf.gov/statistics/2016/nsb20161/#/ (National Science Foundation, 2016).
4. Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. Science faculty's subtle gender biases favor male students. *Proc. Natl Acad. Sci. USA* **109**, 16474–16479 (2012).
5. Reuben, E., Sapienza, P. & Zingales, L. How stereotypes impair women's careers in science. *Proc. Natl Acad. Sci. USA* **111**, 4403–4408 (2014).
6. Steinpreis, R., Anders, K. & Ritzke, D. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study. *Sex Roles* **41**, 509–528 (1999).
7. Williams, W. M. & Ceci, S. J. National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proc. Natl Acad. Sci. USA* **112**, 5360–5365 (2015).

8. Breda, T. & Hillion, M. Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science* **353**, 474–478 (2016).
9. Williams, W. M. & Ceci, S. J. Academic science isn't sexist. *The New York Times* https://www.nytimes.com/2014/11/02/opinion/sunday/academic-science-isnt-sexist.html (31 October 2014).
10. Baumeister, R. F., Vohs, K. D. & Funder, D. C. Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspect. Psychol. Sci.* **2**, 396–403 (2007).
11. Bernstein, R. Scientific community. No sexism in science? Not so fast, critics say. *Science* **346**, 798 (2014).
12. Stewart, A. & Valian, V. *An Inclusive Academy: Achieving Diversity and Excellence* (MIT Press, 2018).
13. Devine, P. G. Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* **56**, 5–18 (1989).
14. Devine, P. G. et al. A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *J. Exp. Soc. Psychol.* **73**, 211–215 (2017).
15. Greenwald, A. G., McGhee, D. E. & Schwarz, J. L. K. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
16. Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L. & Banaji, M. R. Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* **97**, 17–41 (2009).
17. Nosek, B. A. et al. National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proc. Natl Acad. Sci. USA* **106**, 10593–10597 (2009).
18. Rudman, L. A. *Implicit Measures for Social and Personality Psychology* (Sage Publications, 2011).
19. Nosek, B. A., Banaji, M. R. & Greenwald, A. G. Math = male, me = female, therefore math not = me. *J. Pers. Soc. Psychol.* **83**, 44–59 (2002).
20. Miller, D. I., Eagly, A. H. & Linn, M. C. Women's representation in science predicts national gender–science stereotypes: evidence from 66 nations. *J. Educ. Psychol.* **107**, 631–644 (2015).
21. Fazio, R. H. & Olson, M. A. in *Dual-Process Theories of the Social Mind* 155–172 (eds Gawronski, B., Trope, Y. & Sherman, J. W.) (Guilford Press, 2014).
22. Crandall, C. S. & Eshleman, A. A justification–suppression model of the expression and experience of prejudice. *Psychol. Bull.* **129**, 414–446 (2003).
23. Uhlmann, E. L. & Cohen, G. L. "I think it, therefore it's true": effects of self-perceived objectivity on hiring discrimination. *Organ. Behav. Hum. Decis. Process* **104**, 207–223 (2007).
24. Caplar, N., Tacchella, S. & Birrer, S. Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat. Astron.* **1**, 0141 (2017).
25. Hayes, A. F. *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach* (Guilford Press, 2013).
26. Miller, D. I. & Halpern, D. F. The new science of cognitive sex differences. *Trends Cogn. Sci.* **18**, 37–45 (2014).
27. Régner, I. et al. Individual differences in working memory moderate stereotype-threat effects. *Psychol. Sci.* **21**, 1646–1648 (2010).
28. Schmader, T., Johns, M. & Forbes, C. An integrated process model of stereotype threat effects on performance. *Psychol. Rev.* **115**, 336–356 (2008).
29. Valian, V. *Why So Slow? The Advancement of Women* (MIT Press, 1998).
30. Stewart, A. J., Malley, J. E. & Herzog, K. A. Increasing the representation of women faculty in STEM departments: what makes a difference? *J. Women Minor. Sci. Eng.* **22**, 23–47 (2016).
31. Cox, W. T. L. *Multiple Determinants of Prejudicial and Nonprejudicial Behavior*. PhD thesis, Univ. Wisconsin-Madison (2015).
32. Sady, K. & Aamodt, M. G. in *Adverse Impact Analysis: Understanding Data, Statistics, and Risk* (eds Morris, S. B. and Dunleavy, E. M.) 216–238 (Taylor & Francis, 2017).
33. Nosek, B. A., Greenwald, A. G. & Banaji, M. R. Understanding and using the implicit association test: II. Method variables and construct validity. *Pers. Soc. Psychol. Bull.* **31**, 166–180 (2005).
34. Sommers, S. R. On racial diversity and group decision making: identifying multiple effects of racial composition on jury deliberations. *J. Pers. Soc. Psychol.* **90**, 597–612 (2006).
35. Klein, K. J. & Kozlowski, S. W. J. From micro to meso: critical steps in conceptualizing and conducting multilevel research. *Org. Res. Methods* **3**, 211–236 (2000).
36. Chan, D. Functional relations among constructs in the same content domain at different levels of analysis: a typology of composition models. *J. Appl. Psychol.* **83**, 234–246 (1998).
37. Bell, S. T. Deep-level composition variables as predictors of team performance: a meta-analysis. *J. Appl. Psychol.* **92**, 595–615 (2007).
38. Bradley, B. H., Klotz, A. C., Postlethwaite, B. E. & Brown, K. G. Ready to rumble: how team personality composition and task conflict interact to improve performance. *J. Appl. Psychol.* **98**, 385–392 (2013).
39. Devine, D. J. & Philips, J. L. Do smarter teams do better: a meta-analysis of cognitive ability and team performance. *Small Group Res.* **32**, 507–532 (2001)

Acknowledgements

We thank S. Heine, E. Uhlmann, S. Spencer, W. Hall, J. Nezlek and V. Valian for providing valuable input on an earlier version of the manuscript and/or the analyses. Financial support was provided by the Mission pour la Place des Femmes au CNRS (the CNRS Mission for Women's Integration) and the Social Sciences and Humanities Research Council (895-2017-1025; Canada). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This work benefited greatly from technical support provided by the CNRS Mission for Women's Integration. We thank A. Greenwald, N. Sriram and the other members of Project Implicit for efficient technical assistance on the IAT.

Author contributions

I.R., P.H. and C.T.-B. designed the study. I.R., P.H. and C.T.-B. performed the study with technical assistance from A.N.; I.R. and T.S. analysed the data. I.R., C.T.-B. and P.H. supervised the project. I.R., P.H. and T.S. wrote the manuscript. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0686-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to I.R. or P.H.

Peer review information: Primary Handling Editor: Stavroula Kousta.

natureresearch

Corresponding author(s): Isabelle Régner, Pascal Huguet

Last updated by author(s): Jun 28, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The Gender-Science IAT and questionnaire data were gathered on the WEB from Greenwald, Banaji, and Nosek's Project Implicit web-platform.
Data analysis	IBM SPSS statistics 24 with PROCESS macro (Hayes, 2013).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Datafile and codebook are available on on OSF (https://osf.io/umf62/).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Data of the present study were all quantitative data obtained through both explicit (questionnaire) and implicit (Implicit Association Test) measures . Each committee's decision outcomes for years 1 and 2 (i.e., the number of men and women selected and in the candidate pools) were also used to compute an adverse impact (AI) ratio that compares the pass rates of men and women while taking into account the number of men and women in the applicant pool.
Research sample	The study involves human research participants (254 men, 154 women, 6 unspecified) aged from 35 to 64. The age information is not included in our database for ethical reasons related to anonymity : The governing body asked us not to keep this information that, combined with other information (e.g., gender, disciplines, section membership) could help identify the respondents. Participants were all members of the National Committee for Scientific Research, which plays a key role in French science, who were encouraged by the governing body to participate in a study on women's underrepresentation in science. This national committee is a collective body comprised of a general Scientific Board, 10 Institute Scientific Boards, and 40 specialized committees that cover the entire scientific spectrum (from Math and Physical Theories to Political Sciences). As indicated in Figure 1's caption (main text), the committees' structure is ordered by the governing body from STEM fields (numbers from 1 to 20) to life/social sciences (numbers from 21 to 40). Members of these specialized committees (about 20 researchers per committee) meet three times a year typically during 3-4 days.
Sampling strategy	For the period of the study, the population of interest was the CNRS Evaluation Committees which includes 40 committees that represent the whole scientific spectrum from Maths and Physical Theories to Political Sciences. Participation in the study was made available to this entire population. Our final sample includes data from 39 of the 40 available committees. That committee-level data is derived from the scores and responses from 414 committee members (254 men, 154 women, 6 unspecified), which represents a bit more than 50% of the whole population. This participation rate is relatively high considering that participants come from the real world and their decisions would have real consequences for the candidates' actual careers. In contrast to people participating in mock-hiring scenarios, real committee members may be reluctant to report on their attitudes and beliefs about women's underrepresentation in science. And yet, our 50% rate of participation is greater than the 30-35% response rate in prior high profile experimental investigations of gender bias in selection decisions using academic samples.
Data collection	Each participant had free access to a computer throughout the day in a room where they worked alone (15 min in total) during their breaks and other free times. The participation of the committee members was limited to performing the Gender-Science IAT and filling out the questionnaires.
Timing	Participants responses to the questionnaire and IAT were collected from April to June 2010, and the governing body gave us access to each committee's decision outcomes for years 1 and 2 (i.e., the number of men and women selected and in the candidate pools) at the end of 2011.
Data exclusions	Data exclusions were made exclusively on raw data gathered with the Implicit Association Test (IAT) at the individual level (not at the Committee level). As indicated in the Methods and following usual guidelines (Nosek, Greenwald, & Banaji, 2005), IAT scores were disqualified for any of the following criteria suggestive of careless participation: 1) going too fast (<300 ms) on more than 10% of the total test trials, 2) 25% of responses too fast in any one of the critical blocks, 3) 35% too fast in any one of the practice blocks, 4) making more than 30% erroneous responses across the critical blocks, 5) 40% errors in any one of the critical blocks, 6) 40% errors across all of the practice blocks, or 7) 50% errors in any one of the practice blocks. These standards resulted in a disqualification rate of 2.8% .
Non-participation	For the period of the study, the population of interest was the CNRS Evaluation Committees which includes 40 committees that represent the whole scientific spectrum from Maths and Physical Theories to Political Sciences. Participation in the study was made available to this entire population. Our final sample includes data from 39 of the 40 available committees. That committee-level data is derived from the scores and responses from 414 committee members (254 men, 154 women, 6 unspecified), which represents a bit more than 50% of the whole population. This participation rate is relatively high considering that participants come from the real world and their decisions would have real consequences for the candidates' actual careers. In contrast to people participating in mock-hiring scenarios, real committee members may be reluctant to report on their attitudes and beliefs about women's underrepresentation in science. And yet, our 50% rate of participation is greater than the 30-35% response rate in prior high profile experimental investigations of gender bias in selection decisions using academic samples.
Randomization	The present study was a survey, not an experiment. We tested a regressor variable model (i.e., the residual gain score model) that uses Year 1 Adverse Impact as a covariate.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Participants were 414 members (254 men, 154 women, 6 unspecified) of the National Committee for Scientific Research, which plays a key role in French science. To guarantee anonymity, no other information was available.
Recruitment	All the members of the National Committee for Scientific Research, which plays a key role in French science, were encouraged by the governing body to participate in a study on women's underrepresentation in science.
Ethics oversight	The study and potentially related ethical issues were approved by the CNRS governing body at the time of the current study, the Mission for Women's Integration, and the National Committee of Scientific Research (CoNRS General Secretary).

Note that full information on the approval of the study protocol must also be provided in the manuscript.