nature human behaviour

ARTICLES
https://doi.org/10.1038/s41562-020-01033-0

Check for updates

A longer shortlist increases the consideration of female candidates in male-dominant domains

Brian J. Lucas[1]✉, Zachariah Berry[1], Laura M. Giurge[2] and Dolly Chugh[3]

Making it onto the shortlist is often a crucial early step toward professional advancement. For under-represented candidates, one barrier to making the shortlist is the prevalence of informal recruitment practices (for example, colleague recommendations). The current research investigates informal shortlists generated in male-dominant domains (for example, technology executives) and tests a theory-driven intervention to increase the consideration of female candidates. Across ten studies (*N* = 5,741) we asked individuals to generate an informal shortlist of candidates for a male-dominant role and then asked them to extend the list. We consistently found more female candidates in the extended (versus initial) list. This longer shortlist effect occurs because continued response generation promotes divergence from the category prototype (for example, male technology executives). Studies 3 and 4 supported this mechanism, and study 5 tested the effect of shortlist length on selection decisions. This longer shortlist intervention is a low-cost and simple way to support gender equity efforts.

Barriers to gender diversity in the workplace include prejudice, discrimination and differences in social networks[1,2]. Some barriers are referred to as 'second-generation gender bias'[3,4]. These non-deliberate forms of gender bias result from how advancement pathways in organizations are structured. Take, for example, a supervisor who does not consider her top-performing female employees for a leadership position because it involves extensive travel and she assumes female employees do not want to be away from their families. This supervisor may not intend to discriminate, but her beliefs about women's (in)flexibility regarding travel and perception of who 'fits' a leadership role result in a promotion system where women are 'not on the slate'[3]. The current research focuses on one possible source of second-generation gender bias: shortlist generation in the informal recruitment process. We propose and test a low-cost intervention designed to attenuate its effects.

Many professional advancement opportunities (for example, jobs, promotions, skills training and mentorship) are filled through informal recruitment[5,6]. That is, in addition to—or in lieu of—a formal advertisement, candidates are recruited at the suggestion of superiors, colleagues and friends. For instance, a study of 3,100 employers found that 28% did not formally recruit for their most recently filled position[7]. Similarly, research at a large healthcare organization found that 55% of all internal hiring involved a process called 'slotting', whereby managers fill a position by "consulting their mental 'rolodexes'" (p. 853) and choosing a candidate at their discretion[8]. Even opportunities that utilize formal recruitment can involve informal practices. For instance, one survey of new hires revealed that 85% of all jobs were found through some form of networking[9]. Studies of labour markets—such as banking, insurance, phone centres and technology companies—found that between one-third and one-half of candidates in the applicant pool entered through referral rather than direct application[5,10–12]. Overall, this research suggests that informal recruitment is a prevalent pathway to professional advancement.

One consequence of informal recruitment is that it can advantage people who are cognitively accessible in the minds of advancement opportunity gatekeepers (for example, superiors, mentors and industry leaders). When gatekeepers informally recruit by consulting their colleagues or their 'mental rolodexes'[6,8], it can result in a shortlist populated by the cognitively accessible workers that come to mind first[13,14]. While there are meritocratic ways to increase one's visibility, such as positive performance reviews or client feedback[15], informal recruitment may pose an unintended barrier to gender diversity in male-dominant domains because of how cognitive accessibility is influenced by category prototypes and implicit gender stereotypes.

People are more likely to mentally associate a person with a category when that person resembles the category prototype[16]. In a hiring context, candidates who 'fit' the mental representation of a given work role will be more likely to be associated with, remembered for or recommended for this role compared with candidates who do not 'fit' the role[17]. This can lead to unintended consequences in professions where people implicitly associate roles with a specific gender, such as associations between men and the technology industry, the sciences and engineering, many leadership roles, and even action heroes in Hollywood[18–23]. Some widely held implicit gender stereotypes include associations of *female* with *home, family*, *humanities* and *communality*, and associations of *male* with *work, business*, *math, science* and *agency*[24–26]. In many professional domains, these implicit stereotypes inform the perception that male candidates 'fit' the role more appropriately than female candidates. We propose that, for male-dominant roles (for example, technology industry leaders), this perceived 'fit' can lead men to be more cognitively accessible in the minds of gatekeepers and, consequently, more likely to make an informal shortlist compared with equally qualified women.

We suggest a simple intervention to increase the consideration of female candidates in male-dominant domains: make the shortlist longer. This intervention integrates insights from the creativity and brainstorming literatures, which find that generating more (versus fewer) alternatives in a given domain is positively related with divergence from that domain's category prototype[27]. For instance, people

[1]Organizational Behavior Department, Industrial & Labor Relations School, Cornell University, Ithaca, NY, USA. [2]Organisational Behaviour Department, London Business School, London, UK. [3]Management & Organizations Department, Stern School of Business, New York University, New York, NY, USA. ✉e-mail: brianlucas@cornell.edu

Table 1 | Overview of studies

Study	*N*	Preregistered	Role domain	Protocol	Main finding
1a	129	–	Hollywood action hero	Generate initial list of three, then extend the list by three (within-participants design)	More female candidates listed in the extended versus the initial list
1b	87	Yes			
1c	642	Yes			
2a	71	–	Technology industry executive		
2b	194	Yes			
3a	187	–	Role model for a child		
3b	485	Yes			
4	702	Yes	Technology industry executive		
5a	240	–	Hollywood action hero	Generate list of three versus list of six (between-participants design)	More female candidates listed in the longer versus the shorter list
5b	2,166	Yes	Technology industry executive		

N indicates the number of participants in each study's main analysis.

who brainstorm for longer, versus shorter, periods of time generate ideas that are more divergent from the status quo and more creative[28–30]. In professional roles with a dominant gender prototype, making a longer shortlist may increase the likelihood of generating a-prototypical candidates (that is, women for male-dominant roles).

The longer shortlist hypothesis predicts that, in male-dominant domains, making an informal shortlist longer increases the likelihood of listing female candidates. We found support for this hypothesis across ten studies using a variety of shortlist roles (for example, technology industry executive and Hollywood action hero) and participant populations (for example, working adults, and students). See Table 1 for an overview of our studies. The first eight studies (studies 1a–4) used a within-participants design, in which participants generated an initial list of three candidates and then extended the list by adding three more. Across these studies, people tended to list more female candidates in the extended list versus the initial list. The final two studies (5a and 5b) used a between-participants design and found that people listed more female candidates (in both quantity and proportion) when making a longer shortlist (that is, six candidates) versus a shorter shortlist (that is, three candidates). Together, these studies demonstrate that, in male-dominant domains, making an informal shortlist longer increases the likelihood of listing female candidates. Study 3 also found evidence of our proposed prototype divergence mechanism, and studies 5a and 5b found initial evidence of positive downstream consequences of shortlist length on female candidate selection.

Results

Common analysis strategy. Across studies, we conducted the main analyses with Poisson regression because our count data violated normality assumptions (the exceptions to this are noted in the relevant results section). We report effect size as *r* and the ratio of women to men in each time period to aid interpretation. We note that we preregistered our main analyses with repeated-measures *t* tests. However, given the normality assumption violations, non-parametric analyses are more appropriate. For thoroughness, we report the results of the *t* test analyses in the appendix (Supplementary Information, Appendix B) and note in the main text where the parametric and non-parametric tests yield different conclusions. All study preregistrations (1b, 1c, 2b, 3b, 4 and 5b) can be found here: https://osf.io/jb2mq/.

Studies 1a–c. The goal of study 1 was to find initial evidence of the longer shortlist effect in a male-dominant role domain. Participants were 129 university students in study 1a, 87 adults recruited from Amazon's Mechanical Turk (AMT) in study 1b and 642 adults recruited from AMT in study 1c. Participants imagined that they were filmmakers who were asked to generate an informal shortlist of three actors to star in their upcoming action-thriller film (time 1). Then participants were asked to expand the shortlist by adding three more names to the list (time 2). The longer shortlist hypothesis predicts that more women would be listed at time 2 versus time 1. We tested our hypothesis using Poisson regression with participant specified as a panel variable to account for the repeated-measures design. We regressed the number of female candidates listed on time period (1 versus 2). Across studies 1a–c, the number of female candidates tended to be higher at time 2 compared with time 1 (Table 2). This increase was significant in study 1a, non-significant in the predicted direction in study 1b and significant in study 1c.

Studies 2a and 2b. Study 2 tested our hypothesis in an organizational hiring context. Participants were working adults with experience in the technology industry: 71 adults recruited from Prolific Academic (study 2a) and 194 adults recruited from AMT (study 2b). Participants were told about a technology startup in California that was looking for a new chief executive officer (CEO) and were asked to create an informal shortlist of three people who should be interviewed for the role. Then they expanded the list with three additional names. Using the same analytical approach as in the previous study, in study 2a, Poisson regression found significantly more female candidates listed at time 2 compared with time 1 (Table 2). In study 2b, this effect was in the predicted direction but non-significant (we note that this effect was significant in the supplemental *t* test analysis; see Appendix B). An alternative explanation for our effect is that the increase in female candidates at time 2 reflects greater willingness to list less experienced women at time 2. Participants indicated, after completing the task, whether each candidate they listed had executive-level experience (0 for no, 1 for yes). We coded responses for gender and executive experience and found that experience did not differ by gender. The number of candidates with executive-level experience tended to decline as the shortlist got longer for both men (OR 0.88, 95% CI [0.80, 0.98], $Z=-2.38$, $P=0.018$) and women (OR 0.85, 95% CI [0.71, 1.03], $Z=-1.68$, $P=0.093$). The overlapping odds ratio confidence intervals indicate that the decline did not differ in magnitude across gender.

Study 3. The goal of study 3 was to investigate a mechanism underlying the longer shortlist effect. We theorized that making the shortlist longer increases female candidates in male-dominant domains because continued response generation leads responses to diverge from the gender prototype. An alternative explanation is that making the shortlist longer simply increases the tendency to list female

Table 2 | Number of female candidates listed at time 1 versus time 2, studies 1–4

Study	n	Number of female candidates listed				b	95% CI	Z	P	r
		Time 1	95% CI	Time 2	95% CI					
1a	129	0.57 (1:4)	[0.43, 0.72]	0.83 (1:3)	[0.67, 0.99]	.37	[0.07, 0.67]	2.44	0.015	0.21
1b	87	0.18 (1:16)	[0.06, 0.31]	0.29 (1:9)	[0.13, 0.44]	0.45	[−0.18, 1.07]	1.39	0.163	0.15
1c	642	0.28 (1:10)	[0.23, 0.33]	0.36 (1:7)	[0.30, 0.42]	0.24	[0.04, 0.43]	2.41	0.016	0.10
2a	71	0.28 (1:10)	[0.15, 0.42]	0.61 (1:4)	[0.42, 0.79]	0.77	[0.23, 1.30]	2.83	0.005	0.34
2b	194	0.45 (1:6)	[0.34, 0.56]	0.58 (1:4)	[0.46, 0.70]	0.25	[−0.03, 0.53]	1.76	0.079	0.13
3a (parents of boys)	118	0.52 (1:5)	[0.38, 0.65]	0.80 (1:3)	[0.63, 0.97]	0.43	[0.11, 0.75]	2.63	0.009	0.24
3a (parents of girls)	69	1.96 (2:1)	[1.72, 2.19]	1.84 (2:1)	[1.58, 2.10]	−0.12	[−0.46, 0.23]	−0.67	0.505	−0.08
3b (parents of boys)	297	0.55 (1:4)	[0.46, 0.64]	0.78 (1:3)	[0.68, 0.89]	0.36	[0.16, 0.56]	3.50	<0.001	0.20
3b (parents of girls)	188	2.15 (3:1)	[2.02, 2.28]	1.98 (2:1)	[1.84, 2.13]	−0.16	[−0.36, 0.03]	−1.68	0.093	−0.12
4 (baseline)	375	0.47 (1:5)	[0.40, 0.55]	0.61 (1:4)	[0.52, 0.69]	0.24	[0.05, 0.44]	2.43	0.015	0.13
4 (six-name)	327	0.49 (1:5)	[0.40, 0.57]	0.66 (1:4)	[0.57, 0.75]	0.30	[0.10, 0.51]	2.88	0.004	0.16

Parentheses contain women-to-men ratios; in study 3, parents of girls' responses were more evenly distributed and did not follow Poisson distributions, so we analysed them with multilevel linear regression (test statistic *t*); in study 4's six-name-list condition, 'time 1' represents the first half of the list and 'time 2' represents the latter half of the list.

candidates. We tested these competing mechanisms in a domain with same-sex gender prototypes: role models for children (where boys' role models are typically male while girls' role models are typically female). We asked parents to list role models for their male or female child. A simple 'longer list yields female candidates' mechanism predicts that all parents will list more female role models at time 2 versus time 1. However, the prototype divergence mechanism predicts that responses will deviate from the gender prototype: parents of boys will list more female role models at time 2 versus time 1 (deviation from the male prototype) while parents of girls will list fewer female role models at time 2 versus time 1 (deviation from the female prototype).

Parents with a child five years of age or younger were recruited from AMT: 187 parents in study 3a and 485 parents in study 3b. We asked parents to make a list of three role models for their child (time 1), and then expand this list with three additional role models (time 2). Consistent with the longer shortlist hypothesis, for parents of boys, the number of female role models was higher at time 2 compared with time 1 (Table 2). This increase was significant in both studies 3a and 3b. In contrast, for parents of girls, the number of female role models was lower at time 2 compared with time 1. This decrease was non-significant in both studies 3a and 3b (we note that study 3b's effect was significant in the supplemental *t* test analysis, see Appendix B). A comparison of regression coefficient confidence intervals confirmed that, in both studies, the longer shortlist effect was significantly attenuated for parents of girls relative to parents of boys, as indicated by the non-overlapping confidence intervals (Table 2). Study 3 found further evidence of the longer shortlist effect in a different role domain and found evidence consistent with the prototype divergence mechanism.

Study 4. Another alternative explanation for the longer shortlist effect is that it results from generating two different lists (that is, at time 1 and time 2). Participants may perceive this as a form of task switching, which can promote divergent thinking[31]. Study 4 tests this alternative. Participants were 702 adults recruited from AMT with work experience in the technology industry. We randomly assigned participants to either generate a shortlist of technology CEOs by first listing three names and then expanding this list with three additional names (baseline condition), or to generate a single shortlist of six names (the six-name-list condition). Results appear in Table 2. We first looked at the baseline condition. Replicating our prior studies, Poisson regression revealed that participants listed significantly more female candidates at time 2 compared with time 1. Next, we looked at the six-name-list condition. To facilitate analysis comparison across study conditions, we treated the first three names in the six-name-list as 'time 1' and the latter three names as 'time 2'. Contrary to the task switching explanation, participants listed significantly more female candidates in the second half of the list ('time 2') compared with the first half of the list ('time 1'). Consistent with a prototype divergence mechanism, the longer shortlist effect persisted whether the shortlist was generated in one or multiple time periods.

Studies 5a and 5b. Studies 5a and 5b had two goals. First, they tested an implication of the longer shortlist hypothesis. If lengthening a shortlist increases the likelihood of listing female candidates (that is, the longer shortlist effect), then longer shortlists should consist of more female candidates than shorter shortlists. To test this, we randomly assigned participants to generate a three-person shortlist (shorter shortlist condition) or a six-person shortlist (longer shortlist condition). We expected the longer (versus shorter) shortlist to contain more female candidates (in both quantity and proportion). The second goal was to test the effect of shortlist length on subsequent candidate selection. After participants generated their shortlist, we asked them to rank the candidates, with the candidate they most preferred to select ranked as number one. We predicted that the higher number of female candidates in the longer shortlist condition would positively influence the selection of female candidates.

Study 5a participants were 240 university students, and study 5b participants were 2,166 adults with technology industry experience recruited across multiple platforms (see Methods). Participants were randomly assigned to generate a three-person shortlist (list-three condition) or a six-person shortlist (list-six condition). Study 5a used the Hollywood action hero domain, and study 5b used the technology industry executive domain. We first tested the influence of a longer versus shorter shortlist on the number of female candidates listed. Results appear in Table 3. In studies 5a and 5b, Poisson regression revealed significantly more female candidates in the list-six condition compared with the list-three condition. Given that the list-six (versus list-three) condition may have more women simply because it has more slots available, we also compared the proportion of female candidates across conditions and found a similar pattern. In both studies, a Mann–Whitney *U* test revealed a significantly higher proportion of female candidates in the list-six condition compared with the list-three condition (we note that study 5a's

Table 3 | Female candidates listed by condition, studies 5a and 5b

	List-three condition		List-six condition						
	Mean	95% CI	Mean	95% CI	b	95% CI	Z	P	r
Study 5a									
No. of female candidates	0.56 (1:4)	[0.42, 0.70]	1.43 (1:3)	[1.14, 1.71]	0.31	[0.22, 0.41]	6.50	<0.001	0.42
Percentage of female candidates	19%	[0.14, 0.23]	24%	[0.19, 0.29]			−2.18	0.030	0.14
Study 5b									
No. of female candidates	0.51 (1:5)	[0.46, 0.55]	1.18 (1:4)	[1.09, 1.26]	0.28	[0.25, 0.31]	16.79	<0.001	0.36
Percentage of female candidates	17%	[0.16, 0.18]	20%	[0.18, 0.21]			−5.50	<0.001	0.12

Parentheses contain women-to-men ratios; count analyses were conducted with Poisson regressions, and proportion analyses were conducted with Mann–Whitney U tests.

effect was non-significant in the supplemental t test analysis, see Appendix B). Next we tested the influence of the number of female shortlist candidates on candidate selection. First, we conducted a simple mediation model that included condition as the independent variable, number of female candidates as the mediator and female candidate selection as the dependent variable (1 for selected, 0 for not selected). We ran the model with STATA's s.e.m. function using bias-corrected bootstrapped CIs with 10,000 samples. In study 5a, the model revealed a significant indirect effect of condition on female candidate selection via the number of female candidates listed ($b=0.05$, SE$=0.01$, $Z=5.03$, $P<0.001$, 95% CI [0.03, 0.07]); That is, the longer shortlist condition led to more female candidates listed, and this was positively associated with female candidate selection. In study 5b, the model revealed a similar significant indirect effect ($b=0.04$, $SE=0.003$, $Z=13.07$, $P<0.001$, 95% CI [0.04, 0.05]). Finally, we looked at the total effect of shortlist condition on female candidate selection. In study 5a, logistic regression revealed a significant total effect ($Z=2.05$, $P=0.040$, odds ratio 1.28, 95% CI [1.01, 1.62]), such that female candidates were significantly more likely to be selected in the list-six condition ($M=22\%$, 95% CI [0.14, 0.29]) compared with the list-three condition ($M=12\%$, 95% CI [0.06, 0.17]). In study 5b, the effect was in the predicted direction but non-significant ($Z=1.19$, $P=0.235$, odds ratio 1.05, 95% CI [0.97, 1.13]; $M_{\text{list-six}}=17\%$, $M_{\text{list-three}}=15\%$).

Study 5 supported the longer shortlist hypothesis in a between-participants experiment comparing longer versus shorter shortlists. It also demonstrated a positive downstream consequence of shortlist length on female candidate selection: both studies found that shortlist length positively predicted the number of female candidates listed, which positively predicted the number of female candidates selected. The main effect of shortlist length on selection was less consistent. Exploratory analyses of study 5b suggest two competing factors that, together, contributed to that study's non-significant selection effect. First, a longer (versus shorter) shortlist significantly increased the number of female candidates listed, which positively influenced female candidate selection (the indirect effect mentioned above). For example, the number of shortlists that excluded female candidates decreased from 62% in the list-three condition to 42% in the list-six condition, ($Z=9.63$, $P<0.001$, 95% CI [0.17, 0.25]), thus, raising the odds of selecting a female candidate from zero to non-zero in 20% of the lists. However, at the same time, because women were under-represented, increasing the list length, in some cases, added more men than women. This reduces the statistical odds of selecting a woman (for example, a given female candidate has a 33% chance of selection on a three-person list but only a 17% chance of selection on a six-person list). Overall, a longer (versus shorter) shortlist increased the total number of women considered (which positively influenced the odds of female candidate selection) but also, in some cases, added more men than women to the list (which negatively influenced the odds of female candidate selection), producing an overall non-significant effect of list length on selection. These nuanced findings warrant future research and are consistent with research finding that different steps of the advancement process (for example, being shortlisted versus selected) may have different antecedents[32].

Discussion

Ten studies found evidence of a longer shortlist effect: when a participant generated an informal shortlist for a role in a male-dominant domain, more female candidates were included in an extended shortlist (that is, time 2 or the latter half of the list) versus an initial shortlist (that is, time 1 or the first half of the list). We found this effect both when comparing an extended versus an initial shortlist (studies 1–4) and when comparing the second half versus the first half of a single shortlist (study 4). To obtain a more precise effect size estimate, we conducted a mini meta-analysis of this effect across our studies[33] and found an effect size of $r=0.14$ (95% CI [0.10, 0.17]) (for more details, see Supplementary Information, Appendix C). We demonstrated this effect in laboratory (studies 1 and 5) and online (studies 1–4) settings, in samples with industry-relevant knowledge (technology industry, studies 2, 4 and 5) and across diverse shortlist contexts. We proposed that making a longer shortlist increases the consideration of female candidates because continued response generation leads responses to diverge from the gender prototype. Study 3 supported this prototype divergence mechanism by demonstrating that generating more alternatives can increase or decrease the number of female role models on a shortlist, depending on whether the gender prototype is male (increased female role models) or female (decreased female role models), and study 4 ruled out task switching as an alternative mechanism. Finally, study 5 directly tested the effects of shortlist length in a between-participants design and found that generating a longer (versus shorter) shortlist increased the number of female candidates listed and that this positively influenced female candidate selection.

This research contributes to the gender diversity and hiring literatures by identifying informal shortlist construction as a source of unintentional, second-generation gender bias[4]. For professional roles with a male gender prototype—for example, the technology and engineering industries and many leadership positions—a male worker may be more likely to make the shortlist over an equally qualified female colleague simply because he fits the gender prototype and is cognitively accessible to the shortlist generator. A second contribution is that we identify making the shortlist longer as a low-cost, theory-driven intervention to attenuate shortlist construction biases. From the creativity and brainstorming literatures, we import the insight that continued alternative generation promotes category divergence and suggest that prompting gatekeepers to generate a longer informal shortlist can increase the consideration of female candidates. Across studies, making the shortlist longer ($N=3,308$) increased the ratio of women to men from 1 in 5.52 in the initial (or first half of the) list, to 1 in 3.92 in the extended (or latter half of the) list (or from 15% to 20% female candidates). Similarly, study 5 ($N=2{,}406$) found a higher proportion of women in the longer shortlists (20%) compared with the shorter shortlists (17%).

We also contribute to the creativity literature by conceptualizing gender diversity as a 'creative' outcome. Whereas most research on creativity and diversity investigates how diversity influences creativity[34–36], we raise the question of how creativity—and its known antecedents—influences diversity. We find that continued response generation, a demonstrated antecedent of creativity[29,30,37], promotes shortlist gender diversity. We contribute toward a call in the creativity literature for more research on the downstream consequences of creativity[38]. Future research could explore whether diversity is a consequence of other creativity antecedents such as creative leaders or norms for creativity.

Practically, our research provides initial evidence for a simple intervention to help reduce gender bias in the informal shortlist construction process: making a longer shortlist. This intervention has the benefit of simplicity and requiring few organizational resources, particularly when candidates necessitate minimal vetting, as is often the case with informal opportunities such as skills training or mentorship. The intervention also has the flexibility to be implemented as an organizational policy or at the discretion of individual decision-makers. Nonetheless, while our research provides evidence for the effectiveness of this theory-driven intervention in different contexts and with diverse samples, it has yet to be examined in field settings. The absence of field studies is a key limitation of the present research and should be addressed in future studies.

Furthermore, while we found consistent evidence that a longer shortlist increases the inclusion of female candidates, many lists still contained a higher proportion of male candidates and the positive effect of list length on female candidate selection was not consistent in studies 5a and 5b (although the positive indirect effect was consistent). This suggests that organizational leaders should think of the longer shortlist intervention as one of many possible interventions that could be implemented to promote equitable hiring practices. Such a multi-pronged approach is consistent with the systematic nature of inequity and the rarity of effective quick fixes and silver bullets[39,40].

Future research could further test the scope of the longer shortlist effect. Although we found consistent evidence for our hypothesis across various contexts and participant samples, we did observe a wide range of effect sizes ($r=0.10$ to 0.34), suggesting the need for future research to test moderators and boundary conditions; For example, future research should examine the longer shortlist effect in organizations and the possible moderating role of organizational context factors such as job level (entry versus senior position) or the opportunity type (for example, new position, promotion and skills training). Another interesting question is whether the effect is moderated by gender-relevant beliefs or ideologies. For instance, exploratory analyses of study 3's data found that the effect is not moderated by participants' attitudes toward sexism (Supplementary Information, Appendix D), suggesting that the effect occurs, to some extent, outside of awareness.

Future research could also investigate other forms of diversity, such as race/ethnicity, age or expertise. For instance, we began this project also interested in racial/ethnic diversity (see the preregistrations for studies 1 and 2). However, these initial investigations found mixed results, producing, across studies, a non-significant but positive effect of making the shortlist longer on the number of racial/ethnic minority candidates listed ($r=0.02$, 95% CI [−0.01, 0.05], $P=0.123$; Supplementary Information, Appendix E). Exploratory analyses suggest that the effect on racial/ethnic diversity was weaker because the initial shortlists of racially/ethnically diverse participants already exhibited high levels of racial/ethnic diversity, relative to population-level representation, leaving less room for the number of racial/ethnic minority candidates to increase. Future work could explore this and other forms of diversity. Finally, future work should explore factors that predict when making a shortlist longer leads to selection, for example, person factors such as the shortlist generator's demographic diversity and diversity mindset, or process factors such as how the decision is framed[41–43].

We identified informal shortlist generation as a source of gender bias and proposed lengthening the shortlist as one way to attenuate bias. We hope that the longer shortlist intervention helps researchers and practitioners alike to think more creatively about diversity.

Methods

This research was approved by the Institutional Review Board at Cornell University and complies with all relevant ethical regulations. In all studies, participants approved a statement of informed consent before participation. All participants were compensated for their time with either course credit (student participants in studies 1a, 5a and 5b) or a flat fee (all other study participants).

Studies 1a–c. Studies 1a–c tested our hypothesis in the domain of Hollywood action heroes. Across studies, we excluded participants who did not complete the main task or provided nonsense responses (for example, 'good good'), and we excluded participants who failed an attention check that asked them to select a specified number from a list of numbers. These exclusion criteria were applied across all studies and were preregistered in studies 1b, 1c, 2b, 3b, 4 and 5b. Sample size in study 1a was determined by course enrolment. Study 1a recruited 131 university students from an introductory organizational behaviour class ($M_{age}=18.89$ yr, $s.d._{age}=2.16$ yr; 50% male, 50% female); two did not complete the task and zero failed the attention check, leaving 129 participants for analysis. Studies 1b and 1c were conducted online via Amazon's Mechanical Turk (AMT). In these studies, and all subsequent online studies, respondents were required to pass a captcha image check to gain access to the study and the AMT platform restricted repeat participants. Sample size in study 1b was determined by a heuristic of 100 participants for a within-participants test. Study 1b recruited 98 adults ($M_{age}=34.76$ yr, $s.d._{age}=10.27$ yr; 64% male, 36% female); 11 did not complete the task, and 0 failed the attention check, leaving 87 for analysis. For study 1c, we conducted an a priori power analysis with data from study 1b ($d=0.13$, two sided) and determined that 624 participants would yield 90% power to detect the effect. Across our preregistered studies, we aimed for at least 80% power to detect the target effect and increased this power level when our resources allowed. Study 1c recruited 701 adults ($M_{age}=36.31$ yr, $s.d._{age}=11.57$ yr; 54% male, 46% female); 47 did not complete the task and 12 failed the attention check, leaving 642 for analysis. Preregistrations for studies 1b (registered 27 November 2018) and 1c (registered 29 November 2018) are here: https://osf.io/jb2mq/.

Participants completed an informal recruitment task in which they imagined being a filmmaker and needed to generate a shortlist of actors to star in their upcoming film. They were told it would be an action-thriller film that is "packed with action, car chases, and shoot-outs" and they have to cast one lead actor. Participants first generated a shortlist of three names. After completing the shortlist, participants were instructed to expand the shortlist by generating three additional names. We tested our hypothesis by comparing the number of female candidates generated in time 1 versus time 2. To minimize concerns that participants would strategically list actors on the basis of their box office appeal, we told participants that they are talented filmmakers and influential in the field and that their film is likely to be a success no matter who they cast. Female actors were coded as 1 and male actors as 0. All coding was done by two independent coders blind to the shortlist time period. Initial reliability was high ($\alpha=0.99$), and any discrepancies were resolved through discussion. We note here that this study also included another task in which participants listed pairs of actors to play the lead couple in a romantic comedy film and tested the hypothesis that relationship diversity (that is, same-sex versus mixed-sex couples) would be higher at time 2 versus time 1. This hypothesis saw inconsistent support (Supplementary Information, Appendix A).

Study 2. We recruited participants with work experience in the technology industry. The sample size for study 2a was determined heuristically. Study 2a recruited 166 adults through Prolific Academic ($M_{age}=35.81$ yr, $s.d._{age}=10.85$ yr; 63% male, 37% female); 95 were excluded for providing incomplete or unverifiable responses (that is, responses did not appear to be real names and could not be verified through an internet search) and 0 failed the attention check, leaving 71 for analysis. Participants with technology industry work experience were targeted through a pre-existing screener question implemented by Prolific Academic, and we asked participants to report years of technology industry work experience ($M=6.98$ yr, s.d. $=6.77$ yr). The sample size for study 2b was based on an a priori power calculation using data from study 2a (over 90% power to detect half the effect size found in study 2a ($d=0.25$; two-tailed), which we thought seemed high). Study 2b recruited 234 adults through AMT ($M_{age}=32.53$ yr, s.d.$_{age}$ = 7.99 yr; 63% male, 37% female); 37 were excluded for providing incomplete or unverifiable responses and 3 failed the attention check, leaving 194 for analysis. We advertised the study for technology industry workers and asked participants to report technology industry work experience (M = 6.43 yr, s.d. = 6.59 yr). The preregistration for study 2b (registered 17 December 2018) is here: https://osf.io/jb2mq/.

Participants were told to take the role of a corporate strategist who is consulting with a new technology startup in California. They were told that the "company has outgrown its founder and needs an experienced executive to take on the role of CEO." Participants were asked to generate a shortlist of three people the company should consider interviewing. After completing the shortlist, participants were instructed to expand the shortlist by generating three additional names. After the task was complete, we asked participants to indicate each candidate's gender (coded as 0 for male, 1 for female, and 1 for other) and whether that person has prior executive-level experience (0 for no, 1 for yes). Here and across studies, we collapsed the gender categories 'female' and 'other' because there were not enough non-binary responses to meaningfully analyse (for example, two candidates in study 2).

Study 3. We used AMT to recruit parents with a child five years of age or younger. For study 3a, we did not screen by child gender, and if parents had more than one child, they were instructed to choose one of them to focus on in the study. We set our recruitment goal heuristically at 200 responses and received 193 complete surveys (M_{age} = 34.28 yr, s.d.$_{age}$ = 8.44 yr; 58% male, 42% female); 2 did not complete the task and 4 failed the attention check, leaving 187 for analysis: 118 parents of boys and 69 parents of girls. For study 3b, we screened by child gender to account for the oversampling of boys in study 3a. We determined our recruitment goal with an a priori power analysis using data from study 3a (200 parents of boys would provide over 90% power to replicate study 3a's effect on parents of boys (d = 0.33; two-tailed), although we note that a survey programming error led us to oversample parents of boys relative to parents of girls). We received 513 complete surveys (M_{age} = 32.90 yr, s.d.$_{age}$ = 7.46 yr; 52% male, 48% female); 21 participants did not complete the task and 7 failed the attention check, leaving 485 participants: 297 parents of boys and 188 parents of girls. The preregistration for study 3b (registered 14 February 2019) is here: https://osf.io/jb2mq/.

Parents were asked to provide their child's age, gender and first name, the latter of which was piped throughout the survey to personalize the study experience. Parents were told that role models are people that "provide guidance and inspiration as a child develops into an adult" and that a role model can be a famous or public figure or someone they know personally. Then we asked parents to list three role models for their child, aside from themselves. Parents listed the role model's name as well as a website for public figures (for example, Oprah Winfrey) or a description of the person for non-public figures (for example, the child's grandmother). Next, parents were instructed to expand the list with three additional role models. After both lists were created, we showed parents their lists and asked them to indicate each role model's gender. In study 3a there were two gender response options: male and female (coded 0 and 1, respectively). In study 3b there were three gender response options: male, female and other (coded 0, 1 and 1, respectively).

Study 4. Our recruitment goal was 806 participants (403 per condition), providing 80% power to replicate the longer shortlist effect found in pilot data (d = 0.14, two-tailed). We recruited 839 adults from AMT (M_{age} = 33.09 yr, s.d.$_{age}$ = 9.12 yr; 65% male, 35% female). Participants were screened for work experience in the technology industry using the same procedure as in study 2 (M = 6.94 yr, s.d. = 6.47 yr). We excluded 133 participants who did not complete the task or provided unverifiable responses and 4 who failed the attention check, leaving 702 for analysis. The preregistration for study 4 (registered 4 June 2019) is here: https://osf.io/jb2mq/.

In study 4, we used the technology CEO domain from study 2 and randomly assigned participants across two conditions. The baseline condition replicated the procedures of study 2b. Participants generated names for a shortlist of technology CEOs by providing three names at time 1 and three additional names at time 2. To increase the likelihood of obtaining valid responses, we also asked participants to provide a website link at which we could verify the identity of each candidate they listed. As in previous studies, we predicted more female candidates in the extended (versus initial) list. In the six-name-list condition, participants were asked to generate a single shortlist of six names. Thus, in both conditions, participants generated a shortlist of six names. The prototype divergence explanation predicts an increase in female candidates as the shortlist gets longer. To mimic the analyses of the baseline condition, we tested the number of female candidates in the second half of the list compared with the first half of the list. If the effect is driven by perceived task switching, then the number of female candidates should not differ from the first to the second half of the list. Alternatively, the prototype divergence mechanism would predict an increase from the first half to the second half of the list. After generating their shortlists, participants indicated the gender of each candidate (coded 0 for male, 1 for female and 1 for other).

Study 5. For study 5a, we recruited 249 students from an introductory organizational behaviour class (M_{age} = 19.05 yr, s.d.$_{age}$ = 1.61 yr; 42% male, 58% female). The sample size was determined by class enrolment. Seven participants did not complete the task and 2 failed the attention check, leaving 240 for analysis. For study 5b, our recruitment goal was 2,300 participants, providing 95% power to detect study 5a's effect of list length both on female candidates listed (d = 0.19) and on female candidate selection (OR 1.28, two-tailed). We recruited participants with experience in the technology industry by sampling from three platforms: Prolific Academic, AMT and a university-managed participant pool. Similar to previous studies, we expected exclusions, and we collected 3,426 total responses. We excluded 1,221 for not completing the task or providing unverifiable responses and 39 for failing the attention check, leaving 2,166 for analysis (M_{age} = 33.16 yr, s.d.$_{age}$ = 9.98 yr; 64% male, 36% female; 198 from Prolific Academic, 1,765 from AMT and 203 from the university-managed participant pool). We screened for work experience on Prolific Academic using the pre-existing screener question (same as in study 2a; M = 10.20 yr, s.d. = 8.86 yr) and on AMT (M = 7.02 yr, s.d. = 6.80 yr) and the university-managed participant pool (M = 2.73 yr, s.d. = 4.74 yr) using the same protocol as in study 2b. The preregistration for study 5b (registered 21 April 2020) is here: https://osf.io/jb2mq/.

In study 5a, participants completed the action-thriller shortlist generation task used in study 1. In study 5b, participants completed the technology executive shortlist generation task used in studies 2 and 4. Participants were randomly assigned to generate a three-person shortlist (list-three condition) or a six-person shortlist (list-six condition). Then participants ranked the candidates, with the person they would most preferred to select ranked as number one. Participants indicated the gender of each candidate (coded as 0 for, 1 for female and 1 for other).

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

Data availability

All datasets necessary to interpret, replicate and build upon the findings reported in this article can be found here: https://osf.io/jb2mq/

Received: 23 July 2019; Accepted: 9 December 2020;
Published online: 18 January 2021

References

1. Rosette, A. S., Akinola, M. & Ma, A. in *The Oxford Handbook of Workplace Discrimination* vol. 1 (eds A. J. Colella & E. B. King) (Oxford Univ. Press, 2016).
2. Joshi, A., Neely, B., Emrich, C., Griffiths, D. & George, G. Gender research in AMJ: an overview of five decades of empirical research and calls to action. *Acad. Manag. J.* **58**, 1459–1475 (2015).
3. Ibarra, H., Ely, R. J. & Kolb, D. M. Women rising: the unseen barriers. *Harv. Bus. Rev.* **91**, 60–74 (2013).
4. Ely, R. J., Ibarra, H. & Kolb, D. M. Taking gender into account: theory and design for women's leadership development programs. *Acad. Manag. Learn. Edu.* **10**, 474–493 (2011).
5. Fernandez, R. M., Castilla, E. J. & Moore, P. Social capital at work: networks and employment at a phone center. *Am. J. Sociol.* **105**, 1288–1356 (2000).
6. Marsden, P. V. & Gorman, E. H. in *Sourcebook on Labor Markets: Evolving Structures and Processes* (eds I. Berg & A. L. Kalleberg) (Kluwer Academic/Plenum, 2001).
7. Barron, J. M., Bishop, J. & Dunkelberg, W. C. Employer search: the interviewing and hiring of new employees. *Rev. Econ. Stat.* **67**, 43–52 (1985).
8. Keller, J. R. Posting and slotting: how hiring processes shape the quality of hire and compensation in internal labor markets. *Admin. Sci. Q.* **63**, 848–878 (2017).
9. Belli, G. How many jobs are found through networking, really? in *PayScale* (2017).
10. Petersen, T., Saporta, I. & Seidel, M. L. Offering a job: meritocracy and social networks. *Am. J. Sociol.* **106**, 763–816 (2000).
11. Kirnan, J. P., Farley, J. A. & Geisinger, K. F. The relationship between recruiting source, applicant quality, and hire performance: an analysis by sex, ethnicity, and age. *Pers. Psychol.* **42**, 293–308 (1989).
12. Leicht, K. T. & Marx, J. The consequences of informal job finding for men and women. *Acad. Manag. J.* **40**, 967–987 (1997).
13. Srull, T. K. & Wyer, R. S. Category accessibility and social perception: some implications for the study of person memory and interpersonal judgments. *J. Pers. Soc. Psychol.* **38**, 841–856 (1980).
14. Dasgupta, N. & Asgari, S. Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *J. Exp. Soc. Psychol.* **40**, 642–658 (2004).
15. Deutsch, M. *Distributive Justice* (Yale Univ. Press, 1985).
16. Fiske, S. T. & Taylor, S. E. *Social Cognition* (Addison-Wesley, 1991).
17. Hogg, M. A. A social identity theory of leadership. *Pers. Soc. Psychol. Rev.* **5**, 184–200 (2001).
18. Eagly, A. H. & Karau, S. J. Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* **109**, 573–598 (2002).

19. Brescoll, V. L. Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders. *Leadersh. Q.* **27**, 415–428 (2016).
20. Brescoll, V. L. & Uhlmann, E. L. Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychol. Sci.* **19**, 268–275 (2008).
21. Rudman, L. A. & Fairchild, K. Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *J. Pers. Soc. Psychol.* **87**, 157–176 (2004).
22. Abraham, M. Gender-role incongruity and audience-based gender bias: an examination of networking among entrepreneurs. *Admin. Sci. Q.* **65**, 151–180 (2020).
23. Atir, S. & Ferguson, M. J. How gender determines the way we speak about professionals. *Proc. Natl Acad. Sci. U. S. A.* **115**, 7278–7283 (2018).
24. Nosek, B. A. et al. National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc. Natl Acad. Sci. U. S. A.* **106**, 10593–10597 (2009).
25. Nosek, B. A. et al. Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* **18**, 36–88 (2007).
26. Rudman, L. A. & Glick, P. Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* **57**, 743–762 (2001).
27. Mednick, S. A. The associative basis of the creative process. *Psychol. Rev.* **69**, 220–232 (1962).
28. Beaty, R. E. & Silvia, P. J. Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthet. Creat. Arts* **6**, 309–319 (2012).
29. Lucas, B. J. & Nordgren, L. F. People underestimate the value of persistence for creative performance. *J. Pers. Soc. Psychol.* **109**, 232–243 (2015).
30. Lucas, B. J. & Nordgren, L. F. The creative cliff illusion. *Proc. Natl Acad. Sci. U. S. A.* **117**, 19830–19836 (2020).
31. Lu, J. G., Akinola, M. & Mason, M. F. "Switching on" creativity: task switching can increase creativity by reducing cognitive fixation. *Organ. Behav. Hum. Dec.* **139**, 63–75 (2017).
32. Biernat, M. & Fuegen, K. Shifting standards and the evaluation of competence: complexity in gender-based judgment and decision making. *J. Soc. Issues* **57**, 707–724 (2001).
33. Goh, J. X., Hall, J. A. & Rosenthal, R. Mini meta-analysis of your own studies: some arguments on why and a primer on how. *Soc. Pers. Psychol. Compass* **10**, 535–549 (2016).
34. Shalley, C. E. & Gilson, L. L. What leaders need to know: a review of social and contextual factors that can foster or hinder creativity. *Leadersh. Q.* **15**, 33–53 (2004).
35. Shin, S. J., Kim, T. Y., Lee, J. Y. & Bian, L. Cognitive team diversity and individual team member creativity: a cross-level interaction. *Acad. Manag. J.* **55**, 197–212 (2012).
36. Leung, A. K. Y., Maddux, W. W., Galinsky, A. D. & Chiu, C. Y. Multicultural experience enhances creativity: the when and how. *Am. Psychol.* **63**, 169–181 (2008).
37. De Dreu, C. K. W., Baas, M. & Nijstad, B. A. Hedonic tone and activation level in the mood-creativity link: Toward a dual pathway to creativity model. *J. Pers. Soc. Psychol.* **94**, 739–756 (2008).
38. Helzer, E. G. & Kim, S. H. Creativity for workplace well-being. *Acad. Manag. Perspect.* **33**, 134–147 (2019).
39. Chang, E. H. et al. The mixed effects of online diversity training. *Proc. Natl Acad. Sci. U. S. A.* **116**, 7778–7783 (2019).
40. Kalev, A., Dobbin, F. & Kelly, E. Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *Am. Sociol. Rev.* **71**, 589–617 (2006).
41. Hugenberg, K., Bodenhausen, G. V. & McLain, M. Framing discrimination: effects of inclusion versus exclusion mind-sets on stereotypic judgments. *J. Pers. Soc. Psychol.* **91**, 1020–1031 (2006).
42. Apfelbaum, E. P., Norton, M. I. & Sommers, S. R. Racial color blindness: emergence, practice, and implications. *Curr. Dir. Psychol. Sci.* **21**, 205–209 (2012).
43. Chang, E. H., Kirgios, E. L., Rai, A. & Milkman, K. L. The isolated choice effect and its implications for gender diversity in organizations. *Manag. Sci.* **66**, 2752–2761 (2020).

Acknowledgements

The authors thank members of the ExPO Lab for useful feedback. The authors received no specific funding for this work.

Author contributions

B.J.L. and D.C. contributed to the project conception. B.J.L. collected and analysed data. Z.B. and L.M.G. collected and analysed data under the supervision of B.J.L. All authors contributed to the manuscript, and all authors approve the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-01033-0.

Correspondence and requests for materials should be addressed to B.J.L.

Peer review information *Nature Human Behaviour* thanks Molly Carnes, Susan Teubner-Rhodes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor(s): Samantha Antusch, Mary Elizabeth Sutherland and Stavroula Kousta.

natureresearch

Corresponding author(s): Brian Lucas

Last updated by author(s): Sep 7, 2020

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	Surveys were constructed on Qualtrics.com
Data analysis	Data were analyzed with STATA 15.0 for Mac.

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

All datasets necessary to interpret, replicate and build upon the findings reported in this article, including data in Tables 1-3, can be found here: https://osf.io/jb2mq/?view_only=16ff222fd1a94e68b53c1ddbf09bec30

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	All studies utilize quantitative experimental study designs.
Research sample	Cornell University undergraduates were used in Study 1A (Mage = 18.89, SDage = 2.16; 50% male, 50% female). Adults from Amazon's Mechanical Turk were used in Study 1B (Mage = 35.20, SDage = 10.50; 66% male, 34% female) and Study 1C (Mage = 36.34, SDage = 11.66; 53% male, 47% female). Adults with experience in the technology industry were recruited from Prolific Academic in Study 2A (Mage = 32.87, SDage = 8.12; 65% male, 35% female) and from Amazon's Mechanical Turk in Study 2B (Mage = 32.58, SDage = 8.00; 63% male, 37% female). Adults with at least one child were recruited from Amazon's Mechanical Turk in Study 3A (Mage = 34.28, SDage = 8.44; 58% male, 42% female) and Study 3B (Mage = 32.90, SDage = 7.46; 52% male, 48% female). Working adults with experience in the technology industry were recruited from Amazon's Mechanical Turk in Study 4 (Mage = 33.09, SDage = 9.12; 65% male, 35% female). Cornell University undergraduates were used in Study 5a (Mage = 19.05, SDage = 1.61; 42% male, 58% female). Study 5b combined participants from a university-managed participant pool, Prolific Academic, and Amazon's Mechanical Turk (Mage = 33.16, SDage = 9.98; 64% male, 36% female). No studies utilized nationally representative samples.
Sampling strategy	Participants were drawn from convenience samples available at the university or an online recruitment platform. Samples sizes for initial studies were determined by the size of the participant pool or with heuristics such as 100 participants per analysis condition (i.e., Study 1A, Study 1B, Study 2A, Study 3A, Study 5A). Samples sizes for follow-up preregistered studies were determined with a priori power analyses that provided 80% power or higher to detect an effect size obtained in an initial study (i.e., Study 1C, Study 2B, Study 3B, Study 4, Study 5B). Thus, each of our study contexts (movie actors, technology executives, role models) included at least one preregistered test with sufficient power to detect the effect, as previously found in that study context, with at least 80% power.
Data collection	All studies were delivered through computerized surveys built with Qualtrics. In Study 1A the survey was administered in a laboratory setting with one experimenter that was blind to hypothesis (and there was only one condition). In all other studies, the survey was delivered electronically and participants did not interact with any research personnel.
Timing	Data were collected from December of 2018 to July of 2020.
Data exclusions	Participants were excluded from our studies via exclusion criteria that were set a priori and that were the same across all studies: not completing the task (or completing it with responses that did not appear to be real names and could not be verified by an internet search) and/or failing an attention check. Exclusions were as follows: Study 1A, two did not complete the task and zero failed the attention check; Study 1B, 11 did not complete the task and zero failed the attention check; Study 1C, 47 did not complete the task and 12 failed the attention check; Study 2A, 95 did not complete the task, zero failed the attention check; Study 2B, 37 did not complete the task and three failed the attention check; Study 3A, two did not complete the task and four failed the attention check; Study 3B, 21 did not complete the task and seven failed the attention check; Study 4, 133 did not complete the task and four failed the attention check; Study 5A, seven did not complete the task and two failed the attention check; Study 5B, 1221 did not complete the task and 39 failed the attention check.
Non-participation	In Study 1A (the lab study) no participants dropped out. In all subsequent online studies it is difficult to accurately estimate this per se. We required users to pass an image check ("captcha") in order to gain access to the survey and at this stage anywhere from 10%-30% of the sample is screened out, many of those screened out likely being bots. Of those who begin the survey, between 1%-5% drop out at some point and do not finish the survey. No participants declined participation at the stage of informed consent.
Randomization	Studies 1-3 utilized correlational designs that compared participants' earlier responses to their own later responses. Study 4 randomly assigned participants to one of two task conditions. Study 5 randomly assigned participants to generate a list of three or six.

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	see above
Recruitment	All participants were convenience samples. Study 1A and 5A participants were undergraduate students in an organizational behavior class, Study 2A and 5B participants were adults from Prolific Academic, and all other participants were adults from Amazon's Mechanical Turk. This creates the possibility that our results are not representative of the general population. Our approach to minimize the potential for bias was to demonstrate our effect across participants sampled from different platforms (university laboratory, Prolific Academic, Amazon's Mechanical Turk) and within different knowledge domains (movie actors, technology executives, role models).
Ethics oversight	The Institutional Review Board at Cornell University

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