"I'm here for my gender, not my skill": Causal reasoning shapes beliefs about merit in response to DEI initiatives

Aarthi K. Popat (aarthi.popat@yale.edu)
Department of Psychology, 100 College Street
New Haven, CT 06510 USA

Julia A. Leonard (julia.leonard@yale.edu)
Department of Psychology, 100 College Street
New Haven, CT 06510 USA

Abstract

Although well-intentioned diversity, equity, and inclusion (DEI) initiatives aim to increase minority representation in elite groups, they can sometimes backfire by causing candidates to question whether they were selected for merit. Prior work in social psychology suggests that this effect is driven mainly by stereotype threat. Here, we propose a novel cognitive framework: DEI initiatives backfire due to causal inference. Specifically, when candidates hear that they were selected based on a DEI initiative and/or enter a group where they are a minority, they may hypothesize that their selection was based more on their identity and less on their merit. Across two pre-registered experiments manipulating selection messages (DEI vs. merit) and statistical gender representation (represented or under-represented in the selected group), we find evidence in favor of our hypothesis. DEI messages and under-representation independently caused successful candidates to attribute their selection more to their identity and less to their merit but did not directly impact perceptions of competence. A third pre-registered experiment revealed that women selectively rated themselves as less competent in DEI contexts when selection tasks were more difficult. Taken together, this work shows that people make different causal hypotheses about their selection into elite groups based on DEI messages and group composition in conjunction with selection task difficulty and their social identity. Importantly, this work paves the way for designing DEI-based initiatives that license more helpful causal inferences about success to ensure that minority candidates thrive in their positions.

Keywords: causal reasoning; DEI; stereotype threat

Introduction

Calls for more representation on the Supreme Court preceded Justice Sonia Sotomayor's appointment as the first woman of color to ever serve in her role (NPR, 2009). In recent years, cases like hers reflect many institution-implemented diversity, equity, and inclusion (DEI) efforts to rectify gender and race gaps in the United States. These efforts have largely been successful in their goal of increasing minority representation in companies and institutions (Dorsey, 2022; Sekaquaptewa et al., 2019), but there is growing evidence that DEI initiatives have the potential to backfire by lowering candidates' perceptions of their own competence and belonging (Leslie, 2019; Georgeac & Rattan, 2023; Starck, 2021). To this day, even Justice Sotomayor attributes part of her success to her identity (Acevedo, 2023). However, it is unclear exactly *why* and *under what conditions* DEI initiatives backfire.

Prior work in social psychology proposes that DEI messaging instills self-doubt by highlighting people's minoritized identities (e.g., gender, race), thereby raising candidates' concerns about being negatively treated or confirming negative stereotypes related to their identity (Starck et al., 2021; Georgeac & Rattan, 2023; van Veelen et al., 2019; Steele & Aronson, 1995; Casad & Bryant, 2016). Indeed, identity is central to the content and goals of many DEI messages; these efforts often explicitly target women and minorities, who already experience identity-based threat in elite fields (Clance & Imes, 1978; Aronson et al., 1998; Totonchi et al., 2021). However, existing social theories lack a cognitive framework for how people reason about their selection into elite groups across a broad range of identities and inputs. Here, we bridge work from social psychology and cognitive science to propose that causal reasoning shapes harmful responses to DEI-based initiatives.

Humans are 'intuitive scientists' who constantly construct rich causal hypotheses to understand the world around them (Kuhn, 1989; Nisbett & Ross, 1980; Klahr & Dunbar, 1988). When someone is accepted into a selective group (e.g., a university), a viable hypothesis for their admission is their merit in the category that the group selects on (Eisgruber, 1983; Venti & Wise, 1981). However, DEI messaging raises an alternative hypothesis: That identity, instead of or in conjunction with merit, is responsible. In these situations, one might conclude that identity is just a bonus—a supplement to already stellar merit. Alternatively, given that successful learners can discriminate between multiple hypotheses for an event in order to choose the most plausible one (Klahr et al., 1989; Klayman & Ha, 1989; Dougherty & Hunter, 2003), it seems likely that people may privilege one causal hypothesis for their success over another. In other words, successful diverse candidates may feel compelled to up-weight the "identity" hypothesis and down-weight the "merit" hypothesis in response to DEI initiatives.

Causal attributions, especially in the context of reasoning about social groups, are often supported by statistical inference (e.g., Heck et al., 2021). Thus, people's causal hypotheses about their selection into an elite group may also be shaped by the statistical composition of the group. For example, when someone enters a group where they are a minority, they may infer that they were a "diversity hire" in the current political climate—even in the absence of DEI messaging (Hughes, 2013; Estrada et al., 2016; Mathieu, 2023). In the presence of DEI messaging, statistical under-representation could exacerbate identity-based (over merit-based) attributions by indicating to successful candidates that the company or institution did

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

in fact need diversification. By the same logic, the first hires under a *new* DEI policy may draw even stronger identity-based causal attributions for their success given that existing, non-minority group members were selected in the absence of that policy. Since even infants are adept at using statistical evidence to draw rich inferences from sparse data (Xu & Garcia, 2008; Téglás et al., 2011; Goddu et al., 2021; Lombrozo, 2016), it seems plausible that adults harness this same reasoning to inform their beliefs about their acceptance into elite groups.

In short, we propose that causal reasoning shapes people's beliefs about their selection into an elite group in response to both explicit DEI messages and the statistical composition of the group. Specifically, we predict that DEI messages and under-representation independently spur the hypothesis that identity contributed more, and that merit contributed less, to one's admission. That is, when DEI messages and under-representation are combined—as is often the case in DEI contexts—we predict that identity attributions should be the highest and merit attributions should be the lowest. Finally, we hypothesize that the more people perceive their identity as responsible for their success, the less they will view their ability as responsible, even though this trade-off is not required or always accurate.

Importantly, successful candidates' non-meritorious attributions for their success—driven by DEI messaging—may exacerbate negative self-beliefs about their competence. Given prior metanalytic and qualitative literature on detrimental effects of DEI-type initiatives on candidate self-perception (Farrell & Barao, 2023; Leslie, Mayer, & Kravitz, 2014), we hypothesize that the more individuals think they were selected based on gender, not merit, the less competent they will feel at the selection task. Since we predict that both DEI messaging and under-representation will elicit non-meritorious causal attributions for selection, we expect that people's merit-based attributions will predict their ability ratings most strongly in these circumstances. Finally, identity-based individual differences may exacerbate or temper inferences about competence (e.g., Starck et al., 2021; Casad & Bryant, 2016). Specifically, women and minoritized people experience stereotype and identity threat daily, which may heighten identity-based imposter syndrome and skew attributions for success in selective contexts.

The Current Study

Here, we explore how DEI messages about gender parity and the existing gender composition of a selective group impact successful candidates' (1) inferences about the role of their gender and merit in their selection and (2) their reasoning about their own ability. We focus on gender because it is the most common metric on many company DEI assessments in the United States (Chiu, 2022). While, of course, gender is not a binary construct (Hyde et al., 2019), and women hold intersectional identities (e.g., race) that further diversify their experiences in academia and the workforce (Rosette et al., 2011; Thomas et al., 2011), negative stereotypes about women's competence in general are pervasive and develop as early as six-years-old (Bian et al., 2017). By focusing on gender, we can also explore whether women and men draw different causal inferences about the role that their gender and identity played in their selection under a DEI campaign.

Experiment 1 consisted of a 2x2 design crossing selection messaging (DEI- or merit-based) and statistical representation (where the participant's gender is either under-represented or equally represented in an elite group). Participants completed a novel task and were selected into an elite group under the guise of "merit" or "DEI" criteria. Next, participants saw the statistical composition of the group, which was either split evenly by gender or skewed such that the participant held a minority gender identity. Finally, participants separately rated how much they thought their gender and merit contributed to their selection and judged their ability at the novel task. Experiments 2 and 3 were largely similar to Experiment 1, with some methodological changes to elucidate the impact of gender and task difficulty on Experiment 1 effects. All experiments were pre-registered, and data and analyses can be found here.

Experiment 1

Exp. 1 assessed the impact of DEI messaging and gender under-representation on people's attributions for their selection into an elite group as well as their self-perceived ability on the selection task.

Method

We pre-registered a total sample of 300 participants (after exclusions; 157 women, 143 men) based on a power analysis conducted from pilot data showing that we need 300 people to detect a medium effect in an interaction with a power of 0.80. 162 additional participants were excluded from the final sample due to failure of attention and/or comprehension checks and/or failure to complete 75% of trials successfully. Due to the binary nature of our paradigm, participants who held transgender or non-binary gender identities were also excluded.

Participants were randomly assigned to one of four conditions crossing selection messaging and statistical gender representation (see Figure 1): merit/equal ($N = 65$; 28 women), merit/minority ($N = 76$; 43 women), DEI/equal ($N = 86$; 50 women), DEI/minority ($N = 73$; 36 women). Participant age and gender did not differ by condition (age: $F(3) = 0.95, p = 0.42$; gender: $\chi^2(3) = 4.21, p = 0.24$). To control for prior gender stereotypes, we based our paradigm on a novel visual search task called "daxing". To dax, participants had to find a little green character named Dax in a crowded picture. Participants were told that they were being considered for selection by a new algorithm into an elite group: The "Daxing Club". Participants reported their gender, ethnicity, and age before daxing for four trials (see Figure 1). We secretly rigged participants' performance to control for individual differences, such that everyone

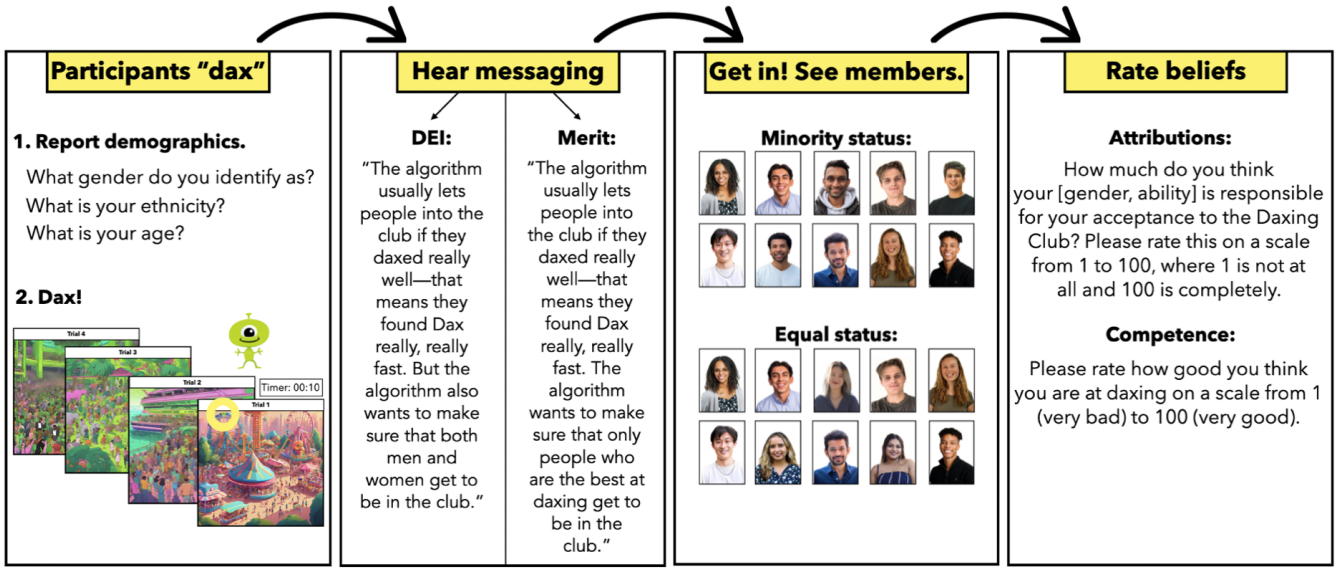

Figure 1: Procedure. Participants reported demographic information to subtly introduce the relevance of gender to the paradigm and then daxed four times. Next, participants heard messaging (DEI- or merit-based), learned of their acceptance to the club, and then "met" the existing members of the club (seeing representation or under-representation). Participants then separately rated their beliefs about how much gender and merit contributed to their selection. Finally, participants rated how competent they think they are daxing.

succeeded in 75% of the trials (i.e., Dax was impossible to find in Trial 3).

After daxing, participants learned of the rationale by which the algorithm would decide whether to admit them (DEI or merit). The merit message told participants that the algorithm only admitted those who were the best (fastest) at daxing, while the DEI message told participants that the algorithm *also* wanted to make sure that both men and women got to be in the club (see Figure 1). Notably, we employed a DEI rationale that also incorporated merit—pragmatically allowing participants to view both identity and merit as equally responsible for their success.

Next, all participants were told they had been accepted to the Daxing Club. Participants subsequently saw the existing club members, whom they were told had been admitted meritoriously before the current algorithm was implemented (see Figure 1: in the *minority* condition, women saw eight men and two women of varied race in the club and men saw eight women and two men; *equal* participants saw five men and five women). Importantly, the specific gender composition of the club was never explicitly highlighted or mentioned and thus had to be gleaned from the visual information alone.

Finally, participants rated their causal attributions for their acceptance into the club based on gender and merit on continuous scales from "not at all" (1) to "completely" (100) on separate pages (order counterbalanced). Importantly, participants' answers did not have to add up to 100. These measures allowed us to test whether people self-imposed a trade-off between gender and merit in their attributions for selection. Finally, we asked participants to rate their own daxing ability on a scale from 1 (very bad) to 100 (very good).

Results and Discussion

Causal attributions. As predicted, DEI messaging and minority status both independently increased participants' gender-based causal attributions for their acceptance and decreased their merit-based causal attributions. A linear model predicting gender attribution with messaging (merit as baseline) and statistical representation (equal as baseline) revealed a positive main effect of DEI messaging ($t(297) = 11.58, p < 0.001$) and under-representation ($t(297) = 3.29, p = 0.001$) on participants' attributions of their selection to their gender. Similarly, a linear model predicting merit attribution with selection messaging (merit as baseline) and statistical representation (equal as baseline) revealed a negative main effect of DEI messaging ($t(297) = -4.79, p < 0.001$) and a negative main effect of under-representation ($t(297) = -3.59, p < 0.001$) on participants' attributions of their selection to their merit. Models with interactions of messaging by statistical representation were not significant when predicting gender attributions ($t(296) = 0.48, p = 0.63$) or merit attributions ($t(296) = -0.75, p = 0.45$). As predicted, we also found that across conditions, gender and merit attributions traded off against each other despite not having to: Higher merit attributions predicted lower gender attributions ($t(295) = -5.15, p < 0.001$).

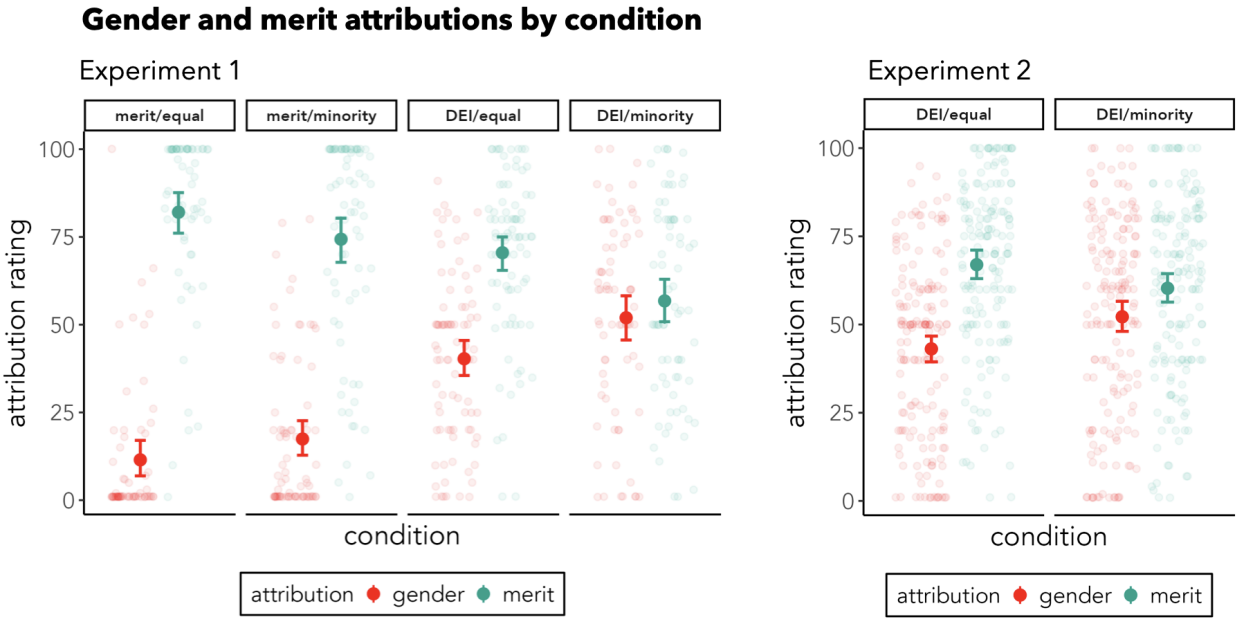

Figure 2: Causal attribution results in Exps. 1 and 2. Across studies and conditions, participants' gender and merit attributions traded off against each other. Notably, participants in DEI/minority condition rated their gender attributions higher and merit attributions lower than any other condition. Effects of statistical representation on causal attributions replicated across exps.

Critically, in pre-registered post-hoc comparison tests, participants in the DEI/minority condition attributed their selection more to gender than participants in any other condition (vs. merit/minority, $t(296) = 8.94, p < 0.001$; vs. merit/equal, $t(296) = 10.07, p < 0.001$; vs. DEI/equal, $t(296) = 3.11, p = 0.011$), and less to merit (vs. merit/minority, $t(296) = -4.09, p < 0.001$; vs. merit/equal, $t(296) = -5.65, p < 0.001$; vs. DEI/equal, $t(296) = -3.29, p = 0.006$; ps all Tukey corrected for multiple comparisons.

Ability judgments. As predicted, we found that, across conditions, people who had higher merit attributions also thought that they were better at daxing ($t(295) = 8.20, p < 0.001$). However, this same relationship was not found for gender attributions ($t(295) = -0.43, p = 0.67$).

Contrary to our predictions, DEI messaging did not interact with gender or merit attributions to predict participants' daxing ability self-ratings (interaction with gender attribution: $t(296) = 0.48, p = 0.63$; interaction with merit attribution: $t(296) = -0.75, p = 0.45$). Statistical representation also did not interact with gender or merit attributions to predict participants' daxing ability self-ratings (interaction with gender attribution: $t(296) = 0.38, p = 0.71$; interaction with merit attribution: $t(296) = -0.37, p = 0.71$). Daxing ability self-ratings also did not differ by condition (ps > 0.05; exploratory analysis). However, exploratory analyses revealed an effect of participant gender: A linear model predicting ability ratings revealed a significant interaction between condition and gender ($t(292) = -2.57, p = 0.01$). Post-hoc exploratory tests reveal that in the DEI/minority condition, women self-rated their daxing ability lower than men in the same condition ($t(64.56) = -3.56, p < 0.001$), as well as women in the merit/equal ($t(58.52) = -2.21, p = 0.031$) and DEI/equal ($t(62.47) = -2.11, p = 0.039$) conditions.

Importantly, we confirmed in exploratory analyses that women in the DEI/minority condition were not slower at daxing than women in these other conditions (vs. merit/equal, $t(60.84) = -0.27, p = 0.78$; vs. DEI/equal, $t(69.25) = 0.47, p = 0.64$), or men in the same condition ($t(65.54) = 1.57, p = 0.12$). Furthermore, exploratory analyses revealed that, across conditions, there was no difference by participant gender on causal attributions of selection to gender ($t(292) = -0.90, t = 0.37$) or merit ($t(292) = -0.15, p = 0.88$). Thus, differences in beliefs about daxing ability by gender in the DEI/minority condition could not be driven by differences in performance or beliefs about how gender and merit played into their selection to the group.

In sum, Exp. 1 revealed that DEI messaging and gender under-representation both independently caused successful candidates to think that their acceptance was due more to their gender and less to their merit than merit-based messaging and

equal gender representation. Participants' causal attributions of gender and merit traded off against each other, even though they did not have to. Furthermore, individuals who thought that their merit greatly contributed to their selection also felt they were better at daxing—but this effect was not amplified or diminished based on any particular DEI message framing or representation. Instead, in exploratory analyses, we found that women selectively felt worse about their daxing ability in the DEI/minority condition. In Exp. 2, we aimed to replicate the main causal attribution effects found in Exp. 1 and the exploratory result of a gender effect on ability ratings.

Experiment 2

Exp. 2 was a pre-registered replication of Exp. 1 with the following two changes. First, we used an easier version of the task (Dax was larger) to avoid excluding a high percentage of participants who couldn't find Dax in all Exp. 1 trials (35%). Second, based on findings from Exp. 1 showing a stronger effect of DEI messaging than representation level on causal attributions, we focused on the DEI/equal and DEI/minority conditions in Exp. 2 to specifically interrogate whether the causal attribution and gender effects replicate under this more subtle statistical manipulation. Focusing on these conditions also allowed us to specifically examine whether gender parity can mitigate backfiring impacts of DEI messages.

Method

We recruited 302 participants (after exclusions; 152 women, 150 men) on Prolific for Exp. 2 based on a power analysis conducted on data from Exp. 1 suggesting that we need a sample size of 300 to detect a medium effect size in planned linear models with a power of 0.8. Due to pre-registered exclusion criteria, 64 additional participants were excluded from the final sample due to failure of attention and/or comprehension checks, failure to complete 75% of trials successfully, and/or holding a transgender or non-binary gender identity. Two additional participants over our pre-registered sample size were recruited due to technical issues with Prolific, and were included in our analysis. The Exp. 2 paradigm was identical to Exp. 1, except Dax was slightly bigger to minimize performance-based exclusions and only included the DEI/equal (N = 149; 69 women) and DEI/minority (N = 153; 83 women) conditions. Participant age and gender did not differ by condition (age: $F = 0.66, p = 0.42$; gender: $\chi^2(1) = 1.89, p = 0.17$).

Results and Discussion

Causal attributions. As predicted, gender under-representation (vs. representation) under DEI messaging caused individuals to attribute their success more to their gender and less to their merit. A linear model predicting gender attribution by statistical representation (equal as baseline) revealed a positive main effect of under-representation ($t(300) = 9.90, p = 0.001$) and a linear model predicting merit attribution revealed a negative main effect of under-representation ($t(300) = -2.74, p = 0.007$). As in Exp. 1, gender- and merit-based attributions for success traded off against each other: Higher merit attributions predicted lower gender attributions across conditions ($t(297) = -10.18, p < 0.001$).

Ability judgments. Contrary to our predictions and our results from Exp. 1, we did not find an effect of participant gender on self-judgments of daxing ability across conditions: Pre-registered t-tests revealed no difference in ability ratings by women in the DEI/minority condition versus men in the DEI/minority condition ($t(151) = -0.16, p = 0.88$) or versus women in the DEI/equal condition ($t(150) = -1.36, p = 0.18$).

One possible explanation for this failed gender effect replication lies in the paradigm itself: We purposefully made daxing easier in Exp. 2 than in Exp. 1. Indeed, participants daxed significantly faster in Exp. 2 than in Exp. 1 ($t(675.59) = -22.79, p < 0.001$), and we excluded far fewer participants in Exp. 2 than in Exp. 1 (Exp. 1 exclusion rate: 35%; Exp. 2 exclusion rate: 17.5%). It is therefore plausible that the task was simply too easy to induce self-doubt in participants, and in particular, women. If women felt more confident in their daxing ability in Exp. 2 than in Exp. 1, their ability beliefs might have been less amenable to our condition manipulation. This is consistent with literature showing that stereotype threat emerges when task difficulty increases (Allison et al., 2017). In Exp. 3, we interrogated this possibility by testing whether women rate their daxing ability lower when the task is harder (as in Exp. 1) versus easier (as in Exp. 2) in the DEI/minority condition.

Experiment 3

In Exp. 3 (pre-registered), we tested whether competence ratings by women in the DEI/minority condition are sensitive to task difficulty. We focused specifically on DEI messages and under-representation because these manipulations gave rise to the gender effect observed in Exp. 1, and most closely approximate real-world DEI contexts. Given prior work suggesting that task difficulty moderates stereotype threat (Barber et al., 2020), we predicted that women in the harder condition would rate their own ability lower than women in the easier condition and men in the harder condition.

Method

We recruited 303 participants (151 women, 152 men) on Prolific for Exp. 3, based on a power analysis using the effect size of competence rating differences between women in the Study 1 DEI/equal and DEI/minority conditions. The power analysis indicated that we would need 75 people per group for 0.9 power. We excluded 86 additional participants due to pre-registered exclusion criteria of failure of attention and/or comprehension checks, failure to complete 75% of trials successfully, and/or holding a transgender or non-binary gender identity. Three additional participants over our pre-registered sample size were recruited due to technical issues with Prolific, and were included in our analysis. The Exp. 3 paradigm used the same DEI/minority condition as in

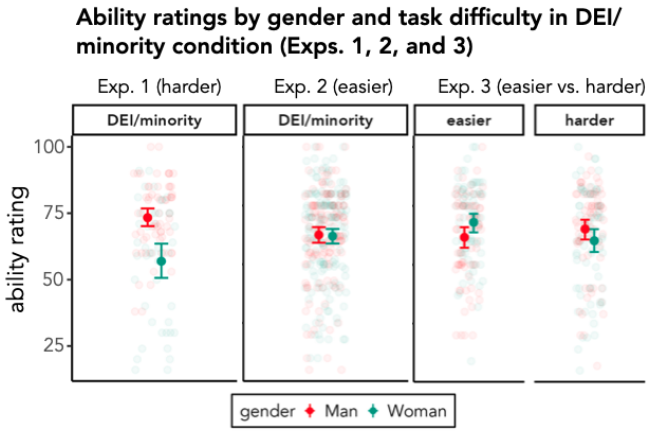

Figure 3: Gender and task difficulty impacted ability ratings in the DEI/minority condition.

Exps. 1 and 2. Half of the participants were in the harder condition (as in Exp. 1; N = 151; 75 women), and half were in the easier condition (as in Exp. 2; N = 151; 76 women).

Results and Discussion

As predicted, a linear model revealed that task difficulty affected women's ratings of their own daxing ability: Women in the harder condition rated their own daxing ability lower than women in the easier condition ($t(149) = -2.47, p = 0.01$). Pre-registered exploratory analyses also revealed an interaction between gender and task difficulty on competence ratings ($t(299) = -2.60, p = 0.01$). In the easier condition, women rated their competence *higher* than men ($t(150) = -2.21, p = 0.03$). However, there was only a trending effect of participant gender within the harder condition: Women in the harder condition rated their competence slightly lower than men in the same condition ($t(149) = -1.52, p = 0.13$).

The results of Exp. 3 revealed that women's daxing ability ratings indeed differ based on task difficulty. Women believed they were worse at daxing in the DEI/minority condition when the task was harder (vs. easier). Although women rated their ability as lower than men in the harder (vs. easier) daxing DEI/minority condition, this effect was not significant, and thus did not conclusively replicate this same gender effect found in Exp. 1. Taken together, Exp. 3 suggests that gender effects in response to DEI contexts are impacted by task difficulty.

Discussion

DEI efforts can unintentionally backfire by causing people to question their own merit. Prior theories in social psychology suggest that this effect is mainly driven by identity-based phenomena like stereotype threat. Here, we present a new theory grounded in cognitive science: DEI efforts backfire due to causal reasoning. Specifically, when people know they were selected under a DEI campaign and/or see that they are a minority in an elite group, they may infer that their selection was based more on their identity and less on their merit. Across experiments, we found evidence in support of this theory. Specifically, we find that the most common DEI initiatives, which involve DEI messages that target under-represented candidates, are also the most likely combination to fuel the inference that selected candidates were accepted for more for their gender and less for their merit when selection tasks are more and less difficult.

Although people made separate, independent judgments about the importance of merit and gender in their selection, we found that, across conditions, people usually privileged one explanation over the other. Under merit-based selection messaging, it makes sense to privilege the merit-based hypothesis. However, when participants heard this same message but entered a group in which they were a minority, they suddenly up-weighted the importance of gender and down-weighted the importance of merit in their selection criteria—even though they were just told that they were selected based on merit. This trade-off between gender and merit attributions became more robust when people heard the DEI selection messages, with people thinking that merit and gender contributed almost equally to their acceptance in the DEI/minority condition. These findings suggest an underlying tendency for people to think that identity and merit-based reasons for selection must trade off against each other. However, another possibility is that people simply assumed that their ratings for merit and gender attributions should add up to 100, since both were queried on sliding 1-100 scales with verbal anchors. It is also possible that the specific wording of our DEI message may have exacerbated this inference. Future work should explore these possibilities.

The more people believed that their success was due to merit in these experiments, the better they thought they were at the selection task across conditions. However, people did not feel automatically worse about their selection task ability when they heard DEI selection messages or were under-represented in the selected group. This indicates that an additional mechanism may link merit-based attributions for selection to ability ratings. Our results from Exp. 3 point to demographic (gender) differences as a potential predictor of lower competence beliefs in DEI contexts: It is possible that identity-based threat *and* causal reasoning work in tandem to cause successful women or minority candidates to doubt their abilities in response to DEI initiatives. Future work should further explore this effect in more threatening domains (e.g., STEM; Spencer et al., 1999).

In academic and professional fields, we constantly wonder how we measure up: Why are we here? Are we good enough? Here, we show that causal reasoning—a cornerstone of human cognition—fuels our identity- and merit-based reasoning about our success in DEI contexts. Understanding this cognitive framework is an essential step toward creating DEI initiatives that signal to targets that their identity does not detract from the merit that led to their success.

Acknowledgments

We thank all members of the Leonard Learning Lab for helpful feedback and discussions. This work was supported by a Jacobs Foundation Research Fellowship to J.A.L.

References

Acevedo, N. (2023, June). *Sotomayor and Jackson slam idea that U.S. is 'colorblind'*. Retrieved 2024-02-01, from https://www.nbcnews.com/news/latino/sotomayor-jackson-write-strong-dissents-affirmative-action-rcna91817

Allison, C., Redhead, E. S., & Chan, W. (2017). Interaction of task difficulty and gender stereotype threat with a spatial orientation task in a virtual nested environment. *Learning and Motivation*, *57*, 22–35.

Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In *Prejudice: The target's perspective* (pp. 83–103). San Diego, CA, US: Academic Press. doi: 10.1016/B978-012679130-3/50039-9

Barber, S. J., Hamel, K., Ketcham, C., Lui, K., & Taylor-Ketcham, N. (2020). The effects of stereotype threat on older adults' walking performance as a function of task difficulty and resource evaluations. *Psychology and aging*, *35*(2), 250.

Bian, L., Leslie, S.-J., & Cimpian, A. (2017, January). Gender stereotypes about intellectual ability emerge early and influence childrenÄôs interests. *Science*, *355*(6323), 389–391. Retrieved 2024-02-01, from https://www.science.org/doi/10.1126/science.aah6524 (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.aah6524

Casad, B. J., & Bryant, W. J. (2016). Addressing Stereotype Threat is Critical to Diversity and Inclusion in Organizational Psychology. *Frontiers in Psychology*, *7*. Retrieved 2024-02-01, from https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00008

Chiu, B. (2022). *For Companies, Diversity, Equity And Inclusion Often Means Gender Diversity.* Retrieved 2024-01-12, from https://www.forbes.com/sites/bonniechiu/2022/09/02/for-companies-diversity-equity-and-inclusion-often-means-gender-diversity/ (Section: ForbesWomen)

Clance, P. R., & Imes, S. A. (1978). The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice*, *15*(3), 241–247. (Place: US Publisher: Division of Psychotherapy (29), American Psychological Association) doi: 10.1037/h0086006

Dorsey, C. (2022, May). A beginnerÄôs guide to diversity, equity, and inclusion typecasting. *Journal of Vascular Surgery*, *75*(5), 1499–1500. Retrieved 2024-02-01, from https://www.jvascsurg.org/article/S0741-5214(21)02743-9/fulltext (Publisher: Elsevier) doi: 10.1016/j.jvs.2021.12.063

Dougherty, M. R. P., & Hunter, J. E. (2003, July). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica*, *113*(3), 263–282. Retrieved 2023-12-20, from https://www.sciencedirect.com/science/article/pii/S0001691803000337 doi: 10.1016/S0001-6918(03)00033-7

Eisgruber, C. (1983). *The Meaning of ÄòMeritÄô in College Admissions.* Retrieved 2024-02-01, from https://paw.princeton.edu/article/meaning-merit-college-admissions

Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Guti©rrez, C. G., … Zavala, M. (2016, September). Improving Underrepresented Minority Student Persistence in STEM. *CBEÄîLife Sciences Education*, *15*(3), es5. Retrieved 2023-12-22, from https://www.lifescied.org/doi/full/10.1187/cbe.16-01-0038 (Publisher: American Society for Cell Biology (lse)) doi: 10.1187/cbe.16-01-0038

Farrell, C., & Barao, L. (2023). Police officer perceptions of diversity efforts: a disconnect between the goals and the methods. *Police Practice and Research*, *24*(2), 216–231.

Georgeac, O. A. M., & Rattan, A. (2023). The business case for diversity backfires: Detrimental effects of organizationsÄô instrumental diversity rhetoric for underrepresented group membersÄô sense of belonging. *Journal of Personality and Social Psychology*, *124*(1), 69–108. (Place: US Publisher: American Psychological Association) doi: 10.1037/pspi0000394

Goddu, M. K., Sullivan, J. N., & Walker, C. M. (2021). Toddlers learn and flexibly apply multiple possibilities. *Child Development*, *92*(6), 2244–2251. Retrieved 2024-02-01, from https://onlinelibrary.wiley.com/doi/abs/10.1111/cdev.13668 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cdev.13668) doi: 10.1111/cdev.13668

Heck, I. A., Kushnir, T., & Kinzler, K. D. (2021). Social sampling: Children track social choices to reason about status hierarchies. *Journal of Experimental Psychology: General*, *150*(8), 1673–1687. (Place: US Publisher: American Psychological Association) doi: 10.1037/xge0001008

Hughes, M. M. (2013). The Intersection of Gender and Minority Status in National Legislatures: The Minority Women Legislative Index. *Legislative Studies Quarterly*, *38*(4), 489–516. Retrieved 2024-02-01, from https://onlinelibrary.wiley.com/doi/abs/10.1111/lsq.12025 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/lsq.12025) doi: 10.1111/lsq.12025

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & Van Anders, S. M. (2019, February). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, *74*(2), 171–193. Retrieved 2024-01-08, from http://doi.apa

.org/getdoi.cfm?doi=10.1037/amp0000307 doi: 10.1037/amp0000307

Klahr, D., & Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. *Cognitive Science*, *12*(1), 1–48. Retrieved 2024-02-01, from https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog1201_1 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog1201_1) doi: 10.1207/s15516709cog1201_1

Klahr, D., Dunbar, K., & Fay, A. L. (1989). DESIGNING GOOD EXPERIMENTS TO TEST BAD HYPOTHESES.

Klayman, J., & Ha, Y.-W. (1989, July). Hypothesis Testing in Rule Discovery: Strategy, Structure, and Content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 596–604. doi: 10.1037/0278-7393.15.4.596

Kuhn, D. (1989). Children and Adults as Intuitive Scientists.

Leslie, L. M. (2019, July). Diversity Initiative Effectiveness: A Typological Theory of Unintended Consequences. *Academy of Management Review*, *44*(3), 538–563. Retrieved 2023-12-20, from https://journals.aom.org/doi/abs/10.5465/amr.2017.0087 (Publisher: Academy of Management) doi: 10.5465/amr.2017.0087

Leslie, L. M., Mayer, D. M., & Kravitz, D. A. (2014). The Stigma of Affirmative Action: A Stereotyping-Based Theory and Meta-Analytic Test of the Consequences for Performance. *The Academy of Management Journal*, *57*(4), 964–989. Retrieved 2023-12-20, from https://www.jstor.org/stable/43589292 (Publisher: Academy of Management)

Lombrozo, T. (2016, October). Explanatory Preferences Shape Learning and Inference. *Trends in Cognitive Sciences*, *20*(10), 748–759. Retrieved 2024-02-01, from https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(16)30105-X (Publisher: Elsevier) doi: 10.1016/j.tics.2016.08.001

Mathieu, F. (2023). *Diversity Hire? Cool, but What About Retention?* (Psy.D., William James College, United States – Massachusetts). Retrieved 2024-02-01, from https://www.proquest.com/docview/2864522921/abstract/D939C5BFF1014799PQ/1 (ISBN: 9798380339100)

Nisbett, R. E., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Retrieved 2024-02-01, from https://philpapers.org/rec/nishis

NPR. (2009, May 26). *Sotomayor: Always looking over my shoulder.* Retrieved from https://www.npr.org/2009/05/26/104538436/sotomayor-always-looking-over-my-shoulder

Rosette, A. S., Ponce de Leon, R., Koval, C. Z., & Harrison, D. A. (2018, January). Intersectionality: Connecting experiences of gender with race at work. *Research in Organizational Behavior*, *38*, 1–22. Retrieved 2024-01-08, from https://www.sciencedirect.com/science/article/pii/S0191308518300121 doi: 10.1016/j.riob.2018.12.002

Sekaquaptewa, D., Takahashi, K., Malley, J., Herzog, K., & Bliss, S. (2019, February). An evidence-based faculty recruitment workshop influences departmental hiring practice perceptions among university faculty. *Equality, Diversity and Inclusion: An International Journal*, *38*. doi: 10.1108/EDI-11-2018-0215

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, *35*(1), 4–28.

Starck, J. G. (2021). *Perpetuating Inequality in Pursuit of Diversity* (Ph.D., Princeton University, United States – New Jersey). Retrieved 2024-02-01, from https://www.proquest.com/docview/2592293390/abstract/E81F463DBCF84C9CPQ/1 (ISBN: 9798460486373)

Starck, J. G., Sinclair, S., & Shelton, J. N. (2021, April). How university diversity rationales inform student preferences and outcomes. *Proceedings of the National Academy of Sciences*, *118*(16), e2013833118. Retrieved 2023-12-20, from https://www.pnas.org/doi/full/10.1073/pnas.2013833118 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.2013833118

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811. (Place: US Publisher: American Psychological Association) doi: 10.1037/0022-3514.69.5.797

Thomas, A. J., Hacker, J. D., & Hoxha, D. (2011, February). *Gendered Racial Identity of Black Young Women.* Retrieved 2024-01-08, from https://link.springer.com/article/10.1007/s11199-011-9939-y

Totonchi, D. A., Perez, T., Lee, Y.-k., Robinson, K. A., & Linnenbrink-Garcia, L. (2021, October). The Role of Stereotype Threat in Ethnically Minoritized StudentsÄô Science Motivation: A Four-Year Longitudinal Study of Achievement and Persistence in STEM. *Contemporary educational psychology*, *67*, 102015. Retrieved 2024-01-12, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8775166/ doi: 10.1016/j.cedpsych.2021.102015

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011, May). Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. *Science*, *332*(6033), 1054–1059. Retrieved 2024-01-07, from https://www.science.org/doi/full/10.1126/science.1196404 (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1196404

van Veelen, R., Derks, B., & Endedijk, M. D. (2019). Double Trouble: How Being Outnumbered and Negatively Stereotyped Threatens Career Outcomes of Women in STEM. *Frontiers in Psychology*, *10*. Retrieved

2024-01-12, from https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00150

Venti, S. F., & Wise, D. A. (1981). *Test scores and self-selection of higher education: College attendance versus college completion* (Tech. Rep.). National Bureau of Economic Research. Retrieved 2024-02-01, from https://www.nber.org/papers/w0709

Xu, F., & Garcia, V. (2008, April). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015. Retrieved 2024-01-07, from https://www.pnas.org/doi/abs/10.1073/pnas.0704450105 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.0704450105