Research: Objective Performance Metrics Are Not Enough to Overcome Gender Bias

Tristan L. BotelhoMabel Abraham

OCTOBER 25, 2017

HBR

Imagine that you are choosing between two similar mutual funds, one managed by Marcus and the other by Tanya. Without additional differentiating information, there is no obvious reason to have a strong preference for one over the other.

Yet in various contexts, such as entrepreneurship and hiring, people often exhibit a preference for men over women when information about an individual's quality (for example, their expected performance) is unavailable or unclear. Even when performance information is available, lab-based research has shown that women still tend to be disadvantaged, compared with men of equal quality. This double standard means women must outperform men to be evaluated similarly.

But how pervasive is this problem, exactly? We wanted to test the extent to which gender-based double standards are present in a competitive context, where evaluators have access to objective performance information and are motivated to evaluate candidates based on quality alone. In this research, we studied investment professionals who are arguably disincentivized to incorporate gender, or any characteristic not directly tied to quality, in their evaluations of investment opportunities. As one industry insider told us, "I am in the business of making money, I don't care who you are, as long as you can help me in that goal."

We collected data from a private online knowledge-sharing platform where buy-side (hedge fund, mutual fund) investment professionals (analysts, portfolio managers) share their investment recommendations to buy and sell a given security, along with the analysis supporting their position, with other members of this platform. We analyzed 3,520 recommendations by 1,550 recommenders of buy and short-sell recommendations for stocks listed on a U.S. exchange over a six-year period (2008 to 2013).

On the platform, members evaluate recommendations across two stages. In the first stage, the "attention stage," they see recommendations and click on ones they want to learn more about. They see only a limited amount of information at this stage: the stock being recommended, the position (buy versus short), the market return of the recommendation, the recommender's name, and the name of the recommender's employer. If they want to access the detailed analysis that supports the recommendation, they click through. In this second stage, the "feedback stage," they see the analysis for those recommendations that they clicked on. Here, they can anonymously rate, on a five-star scale, the recommendation's quality and future performance and publicly leave comments.

The gender of the investment professional recommending a stock is not explicitly stated during either stage of the evaluation process. The only way to infer whether a recommendation was submitted by a man or women is from their first name. As one investment professional told us, "When I see a name like Mary, it sticks out because there are so few [women] in this industry." To parallel this gender inference, we used an algorithm developed by IBM — the IBM InfoSphere Global Name Management Tool — to score the first name of each investment professional, to determine how likely it was that a given member of our sample was female. We were able to gather the self-disclosed gender for about half of our sample, as some members did disclose their gender. (If they did, it's visible on their profile, though it does not appear alongside their recommendations.)

Given the general underrepresentation of women among investment professionals, especially in hedge funds, it was not surprising that only about 4.2% of recommendations were submitted by a person with a name more likely to be associated with a woman than a man. We found no notable gender differences in performance or in investors' backgrounds (though members in our sample with more-female-sounding names were more likely to have attended a top-ranked undergraduate university).

We first examined whether the number of clicks a recommendation received during our study was correlated with the recommender's inferred gender. We controlled for observable characteristics, the most important being recommendation performance, which we calculated as the recommendation's return relative to the S&P 500, focusing on the week following a recommendation being submitted.

We found that a recommendation submitted by someone with a very female name, like "Mary," received approximately 25% fewer clicks overall than a recommendation submitted by someone with a very male name, such as "Matthew." Further, we found that this penalty was largest when evaluators faced high search costs (in other words, when they had more recent recommendations to choose from).

Most surprisingly, this difference remained even when we excluded women from the sample. Focusing our sample on those members who self-disclosed their gender as male, we found that men with more-female-sounding names, such as "Kelly," received fewer clicks. This helps address an important alternative explanation that characteristics (real or imagined) associated with gender, such as personality, could be driving our results. Because investment professionals with names more likely to be considered female do not differ in terms of performance and other observable characteristics, our analysis provides strong evidence that evaluators are incorporating gender into their assessments.

Our results also demonstrated that women had to substantially outperform the average investment professional in order to receive the same attention as their male counterparts. This gender difference disappeared only for the highest-performing (top 10%) recommendations, where gender no longer predicted the likelihood that a recommendation was clicked on.

However, there is a silver lining: In our sample, women were not penalized at the second stage of the evaluation process. We found no gender difference in the ratings or number of comments recommendations received. So while recommendations by individuals with more-female-sounding names received fewer clicks, they were later evaluated similarly to those by men. This is consistent with previous research that has shown people exhibit greater reliance on status signals, such as gender, when there is greater uncertainty about quality. When evaluators had more information about a recommendation (from the detailed analysis) there was less uncertainty about quality, and they were less likely to display any gender-related preference.

Performance is often thought of as the ultimate equalizer, serving as the lynchpin of meritocratic systems. Our findings show that the mere availability of performance information is insufficient for eliminating gender bias — even in a context where decision makers are extremely performance-minded. Rather, our study shows that it's only when evaluators have access to more-comprehensive information with which to base their assessments that bias against women is reduced. This suggests that when women — and probably other underrepresented groups, too — are being evaluated, such

as for a job, they may benefit from presenting more information about their experience and performance, to minimize ambiguity about their quality.

Additionally, our research suggests that while policy interventions, such as providing opportunities for women and minorities in investment management, help redress skewed distributions of the workforce, women still may face unfair assessments once in the field.

There are a few considerations for organizational leaders to take into account. For example, our results suggest that search costs in evaluations (having many options to consider) intensifies gender bias. Therefore, managers should be tasked with evaluating smaller subsets of a larger candidate pool. By capping the number of options, evaluators are better able to spend more time on each candidate, thus reducing the likelihood that bias will creep in. In cases where this is difficult, organizations may redact gender-identifying information about candidates, such as their name. The effectiveness of this approach is limited, since other information about a candidate may signal gender, such as slower upward career mobility for a female candidate.

Finally, our results suggest that offering more pertinent information to evaluators should help decrease gender bias in the evaluation process. The platform under study is a great example of allowing open access to anyone in an industry to present their skills. Organizations could incorporate similar digital platforms internally that allow employees to exhibit their skills more broadly.