Why Being the Last Interview of the Day Could Crush Your Chances

Published : February 13, 2013 in Knowledge@Wharton

Sorry, grad school applicants. According to new Wharton research, not only must prospective students or job seekers compete against a crowded field of equally appealing candidates, but they also must shine when compared to the randomly selected cluster of applicants who have interviews scheduled on the same day.

Like gamblers who swear that a run of red numbers at the roulette table means it's time to bet on black, individuals who are tasked with breaking up a series of decisions over a number of days don't always take the long view when making their judgments. As shown in a research paper co-authored by Wharton operations and information management professor Uri Simonsohn, those decisions are affected not only by the expected overall distribution of results but also by the results seen in a single day's small, unrepresentative sample.

In "Daily Horizons: Evidence of Narrow Bracketing in Judgment from 10 years of MBA-admission Interviews" Psychological Science, Simonsohn and Harvard University professor Francesca Gino used MBA admissions data from a university (that was neither Wharton or Harvard) to study what happened to applicants' scores when they were interviewing at the end of a day and after a series of strong -- or a series of weak -- candidates. ," recently published in

Later in the Day, Lower in the Rankings

Their theory was that a phenomenon called "narrow bracketing" was affecting how those late-day candidates were being judged. Put simply, narrow bracketing is when an individual makes a decision without taking into account the consequences of many similar choices. At the roulette wheel, a gambler who knows that the wheel's odds of turning up red or black are 50/50 will look at the day's results -- which are often displayed by the casino -- and predict a run of a certain color, even though a subset of a croupier's spins is not necessarily representative of the expected overall distribution. Simonsohn and Gino posited that a similar effect happens in the business world, too, when professionals are faced with spreading a long string of similar decisions over multiple days.

On a five-point scale, with five being the best possible score, a similarly qualified applicant who interviewed on the tail end of his top-scoring competition got lower scores overall than what he or she would have otherwise received. Conversely, those who interviewed after a group of weaker competitors got better than expected evaluations. The data covered more than 9,000 interviews done by 31 interviewers, none of whom were alumni.

"If [an interviewer] interviewed four people, and all four have been good, they will think the fifth person is less likely to be good," Simonsohn notes. "Of course, we don't get to see their beliefs [about a candidate], but we get to see how they evaluate the candidate. We wanted to know if they give a lower evaluation [to that fifth person], controlling for everything we know about the person they're talking to. It turned out they do get lower ratings."

The hypothesis, Simonsohn says, is that after giving the first four applicants high ratings, an interviewer may be reluctant to do the same for a fifth candidate if he knows that only a certain percentage of individuals are accepted into a program, or that only some will move to the next stage of a selection process.

"For instance, an interviewer who expects to evaluate positively about 50% of applicants in a pool may be reluctant to evaluate positively many more or fewer than 50% of applicants on any given day. An applicant who happens to interview on a day when several others have already received a positive evaluation would, therefore, be at a disadvantage," Simonsohn and Gino write. By applying the expected overall result of a series of decisions -- in this case, knowing the percentage of candidates accepted into a graduate school program -- to the subset of decisions being made in a particular day, the interviewers are exhibiting narrow bracketing behavior.

"These arbitrarily created subsets should have no influence on experts' judgments," Simonsohn and Gino add. "While the merit of an MBA applicant may partially depend on the pool of applicants that year, it should not depend on the few others randomly interviewed that day."

This phenomenon is not just confined to the academic admissions world, Simonsohn says. He imagines a similar dynamic playing out whenever individuals are spreading similar decisions out over multiple days, including taking loan applications at a bank or interviewing candidates for a job. (While it is less likely to occur when the process gets down to choosing a single hire, it could come into play in an earlier round that reduces the size of the candidate pool.)

"In any setting where people have to make a large set of judgments that is broken down into a small set on the same day, you might see the same thing," he notes.

The Reason behind the Rankings

Simonsohn was able to observe the narrow bracketing phenomenon thanks to the wealth of data both on the MBA candidates (including their GMAT scores) and the interviewers' overall impressions of them, including a number of sub-scores on specific areas (including communication skills, ability to work on a team and interest in the school). The data didn't, however, point to a definitive answer for why this is happening.

The effect very well could be an unconscious one, Simonsohn says, or "it could be very conscious. It could be an agency thing. It could be you don't want your supervisors to think you're doing a bad job when they see a bunch of [candidates rated as] fives in a row."

What the research was able to rule out was the effect of seeing a genuinely less-qualified candidate toward the end of the day. Simonsohn notes that he and Gino did an analysis trying to predict the GMAT scores and experiences of the late-in-the-day, lower-rated candidates based on the interviewer's scores. "We couldn't do it," he says "If that last link really is weaker, you should be able to see evidence of that, and that didn't happen."

The paper was also able to rule out a contrast effect -- in this case, judging an applicant based on the person or persons who were interviewed before him or her, noting that there were no significant differences in the sub-scores that rated candidates on certain attributes. "For example, the contrast between an eloquent applicant and an inarticulate one seen back to back should be starker than that between applicants who differ in their overall strength aggregated across a broad range of attributes," the researchers write.

"The opposite prediction follows from the narrow bracketing account," they continue. "Because interviewers are unlikely to be concerned about keeping a balanced distribution of each sub-score, and they may even have difficulty remembering the sub-scores they gave to previous applicants, sub-scores should more weakly, if at all, be influenced by previous sub-scores."

What Interviewers Can Do

For interviewees, Simonsohn says his findings aren't going to be of much strategic help. "There's no magic in this for the user," he notes. "You can't see who you're competing against and often can't control the timing of your interview.... When the candidates are spread out over weeks and weeks, your competition is the entire applicant pool and not a subset of that. But in reality, your competition is drawn from two pools -- everyone and the other applicants who get interviewed that day."

The effect can be seen even in less-formal daily subsets, Simonsohn and Gino write. "A similar bias may occur when people conduct larger sets of evaluations and generate subsets spontaneously in their minds. Imagine, for example, a judge who must make dozens of judgments a day. Given that people underestimate the presence of streaks in random sequences ... the judge may be disproportionately reluctant to evaluate four, five or six people in a row in too similar a fashion, even though that 'subset' was formed post-hoc."

But companies or universities may be able to control for the narrow bracketing effect in low-cost, low-risk ways, Simonsohn says. His suggestion would be to have interviewers enter each applicant's scores into a spreadsheet or database program that would help them monitor the results of their interviews over time and keep focus off that day's crop of candidates.

"A spreadsheet keeping tabs on the entire interview process can visually present the distribution of your interview scores, and those scores won't jump out at you as much as several interviews in a row," Simonsohn notes. "It's not very sexy, but it's a low-tech solution and it's low risk."

Simonsohn and Gino's next step in their research is to test their proposed solution in a laboratory setting to see if it has an impact on the narrow bracketing effect. "[Hopefully] it really reframes the bias from the short term to the long term," he says.