Chapter 5
Overcoming the Small-N Problem

Iris Bohnet, *Harvard University*

Siri Chilazi, *Harvard University*

SUMMARY

Small samples negatively affect the quality of the information we use when making group-based estimates. Small samples have higher variability than large samples, so data about a handful of female and minority leaders are less informative than the data about the large cohort of their male and white counterparts. While reliance on group-level characteristics, i.e., stereotypes, is in itself hotly debated, the problem of stereotyping is compounded by the data being less accurate and less reliable the smaller the group is. Simply put, if you want to learn about the typical attributes of, say, millennials, a sample of 10,000 will yield more useful information than a sample of 100. In addition, relative to majorities, minorities are more likely to be subject to tokenism, and additional scrutiny, in numerically skewed groups where they make up only a small proportion of the group. People are unlikely to correct for small-N statistics and often erroneously consider small samples to be equally representative of the underlying population as large samples. Obviously, increasing the sample size would solve the small-N problem. As this is not always possible, another way to counteract the threat of inaccurate stereotypes is to increase the availability of role models by making visible individuals from underrepresented groups who are representative of the group as a whole. Additionally, changes in decision processes that decrease the impact of stereotypes on people's judgments by focusing attention on individual-level data rather than group-level characteristics are likely to improve diversity because differences in sample size no longer matter.

KEY FINDINGS

Problems

Group size

- Smaller samples are less informative than larger ones, leading to less accurate, more variable, and less useful inferences. Assessments based on group averages are more likely to be inaccurate.
- Relative to members of larger groups, members of smaller groups are more likely to stand out, receive more scrutiny, and feel pressure to assimilate (tokenism).

Perception of group size

- Smaller samples tend to be taken as equally representative of the population as larger samples (the representativeness heuristic).
- People tend to mistake easily retrievable and salient examples for frequent occurrences (the availability heuristic). For example, a person who remembers that Sheryl Sandberg is the COO of Facebook might estimate the fraction of female corporate leaders to be much higher than it actually is.[1]

1 Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982).

Relevance of group size for decision-making

- A host of situational factors including stress, evaluation procedures, and accountability mechanisms can affect the relative importance of group stereotypes (whose accuracy depends on group size) and individual characteristics (which are independent of group size) in the decision process.
- One of the consequences is stereotype threat, which can undermine the efficacy of stereotyped individuals.[2]

Solutions

Group size

- Make the sample bigger by increasing the representation of underrepresented individuals, such as by setting goals, targets, or quotas.

Perception of group size

- Make a greater number of representative examples (role models) of underrepresented groups more salient and visible, such as by increasing their inclusion in public displays (conference panels; media and movies; public portraits and art; etc.). Over time, this can counteract stereotype threat and even change societal stereotypes.

Relevance of group size for decision making

- Focus attention on relevant individual-level information to decrease the impact of group-based assessments, such as by changing evaluation procedures (joint and simultaneous evaluation instead of separate and sequential evaluation).

A multinational company we are working with is trying to motivate managers to hire and promote more women and people of color by encouraging them to "take more risks." While describing members of smaller groups as intrinsically riskier than members of larger groups might not be the best strategy to advance diversity—the company's ultimate goal—there is some truth to the statement. Women and people of color hold a small fraction of leadership positions in this company, as they do in the vast majority of organizations. Such small samples are inherently more variable than large samples and less likely to be representative of the population distribution, thus yielding less useful information for people who want to learn from them.[3] This makes hiring women or people of color seem riskier. When asked whether women or people of color lead differently than men, for example, people's assessments can be skewed because their answers compare a very small sample of political and business leaders who identify as women and people of color with a much larger sample of white and male leaders. In addition, the

2 Stereotype threat refers to situations where members of a stereotyped group are concerned about being judged in light of the stereotype, which can undermine their performance and aspirations. For example, women are stereotyped to have lower math ability than men, which can result in women performing worse on math tests than their ability would predict, especially if their gender identity is made salient. Nalini Ambady et. al., "Stereotype Susceptibility in Children: Effects of Identity Activation on Quantitative Performance," *Psychological Science* 12, no. 5 (2001): 385–390.

3 Iris Bohnet and Farzad Saidi, "Informational Inequity Aversion and Performance," *Journal of Economic Behavior and Organization* 159 (2019): 181–191.

sample itself may be biased because the very few individuals who have made it to the top are likely not representative of the average individual of a given group.

The "small-N problem" affects numerically underrepresented groups—the "small N members"—in at least three ways: they have fewer and thus less useful role models available to learn from; they are more likely to be taken for tokens; and when being evaluated for hiring or promotion, they are confronted with managers who might be affected by inaccurate stereotypes. But the "small-N problem" also affects members of numerical majority groups, whose decision-making can be unduly influenced by unconscious biases. In order to give everyone an equal shot at success in the workplace, it is essential to overcome the challenges posed by small and biased samples and make unbiased decisions regardless of representation. So how can we make it happen and level the playing field?

The Problems of Small Samples

Decision makers tend to rely on group-based assessments when person-specific information is limited but group-level information is available.[4] Such group-based assessments have been widely documented in labor, housing, and credit markets.[5] Using group stereotypes, however, has direct negative effects leading to differences in pay[6] and opportunity[7] between the advantaged and the discriminated groups. It also has secondary negative effects leading members of the discriminated group to decrease their effort in response to anticipated lower returns to effort, which, in the end, can induce a vicious cycle whereby the individuals who were discriminated against perform worse than they would have in the absence of discrimination.[8] For example, if there is a gender gap in promotion that causes equally qualified women to be promoted less often than men, women might adjust their effort by working less hard, thus confirming the prevailing—but inaccurate—beliefs about their ability. Members of smaller groups are further disadvantaged relative to members of larger groups by tokenism, which magnifies differences and makes minority members subject to additional scrutiny.[9] If a hockey team has 20 different countries represented, nationality is unlikely to be a salient issue; but if only two players come from a different country than the rest of the players, they are likely to stand out.[10]

4 Economists refer to this as "statistical discrimination." People often rely on group-level characteristics in situations where individual characteristics are hard to observe. Edmund Phelps, "The Statistical Theory of Racism and Sexism," *American Economic Review* 62, no. 4 (1972): 659–61; Kenneth J. Arrow, "The Theory of Discrimination," in O. Ashenfelter and A. Rees (eds.), *Discrimination in Labor Markets*, (Princeton, NJ: Princeton University Press, 1973).

5 Marianne Bertrand and Esther Duflo, "Review on Field Experiments on Discrimination," in *Handbook of Field Experiments*, eds. Abhijit Banerjee and Esther Duflo (North Holland, U.K.: Elsevier Science, 2017).

6 Roland Fryer, Devah Pager, and Jörg L. Spenkuch, "Racial Disparities in Job Finding and Offered Wages," *The Journal of Law & Economics* 56, no. 3 (2013): 633-689.

7 Lincoln Quillian, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen, "Meta-Analysis of Field Experiments Shows no Change in Racial Discrimination in Hiring over Time," *Proceedings of the National Academy of Sciences* 114, no. 41 (2017): 10870–10875.

8 Iris Bohnet, Ashley Craig, and Clémentine van Effenterre, "Overcoming Gender Differences in Interview Ratings," Harvard Kennedy School Working Paper, Harvard University, Cambridge, MA, 2019 (Microsoft Word file).

9 Rosabeth M. Kanter, "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women," *American Journal of Sociology* 82, no. 5 (1977): 965–990.

10 Katherine W. Phillips and Damon Phillips, "Nationality Heterogeneity, Performance, and Blau's Paradox: The Case of NHL Hockey Teams, 1988-1998," Paper presented at the Academy of Management conference, New Orleans, LA, August 2004.

Small numbers present a cognitive challenge because of how people perceive them. People do not accurately correct for sample size, believing that smaller samples are as representative of the underlying population as larger samples. Furthermore, the availability heuristic leads human minds to make decisions based on the most easily accessible information—in other words, salient and readily available examples. Americans tend to think of the elderly when asked who lives in Florida even though more than 80 percent of Floridians are younger than 65.[11] Indeed, stereotypes are often based on readily available examples rather than actual prevalence, such that human minds overweight a group's most outstanding types in determining their average characteristics.[12]

Finally, the small-N problem can be amplified by situational factors, such as the design of the decision process.[13] Hiring, promotion, and electoral procedures often focus on very few individuals—sometimes just one—and evaluate candidates separately and sequentially. However, when people evaluate one person at a time and do not make joint assessments, they are more likely to rely on stereotypes instead of individual-level information to make decisions.[14] For example, in first-past-the-post electoral systems common in the United States, the United Kingdom, Canada, and India, voters indicate their preferred candidate on a ballot, and the candidate with the most votes wins. This winner-takes-all, single-member district system stands in contrast to proportional representation electoral systems, which are designed to make the proportion of seats awarded to candidates and parties reflect the proportion of votes they receive as closely as possible. It turns out that proportional representation elections are twice as likely to propel women into office than winner-takes-all elections, and women are even more likely to get elected in multimember constituencies where voters evaluate candidates jointly and make multiple simultaneous decisions by voting for multiple candidates.[15]

Solutions to Overcome the Small-N Problem

The most obvious solution to the small-N problem is to make the sample bigger by increasing the numbers of traditionally underrepresented individuals. In India, this was done successfully through political quotas that randomly assigned a third of village chief positions to women, which caused the share of women in local government to increase from 5 percent in 1993 to 40 percent in 2005.[16] The United Kingdom saw similar success when the government introduced non-binding targets that encouraged companies to diversify their boards of directors. Supported by peer pressure among large companies and senior executive search firms, as well as a research-based publicity campaign, the targets

11 Pedro Bordalo et. al., "Stereotypes," *The Quarterly Journal of Economics* 131, no. 4 (2016): 1753–1794.

12 Devah Pager and Diana Karafin, "Bayesian Bigot? Statistical Discrimination, Stereotypes, and Employer Decision Making," *The Annals of the American Academy of Political and Social Science*, 621, no. 1 (2009): 70–93.

13 Iris Bohnet, *What Works: Gender Equality by Design* (Cambridge, MA: The Belknap Press of Harvard University Press, 2016).

14 Iris Bohnet, Alexandra van Geen, and Max Bazerman, "When Performance Trumps Gender Bias: Joint Versus Separate Evaluation," *Management Science* 62, no. 5 (2016): 1225–1234; Edward H. Chang et al., "The Isolated Choice Effect and Its Implications for Gender Diversity in Organizations," Forthcoming in *Management Science* (n.d.).

15 Pippa Norris, "The Impact of Electoral Reform on Women's Representation," *Acta Politica*, 41, no. 2 (2006): 197–213.

16 Lori Beaman et. al., "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124, no. 4 (2009): 1497–1540.

helped to increase the proportion of women on the corporate boards of FTSE 100 companies from 12.5 percent in 2011 to more than 30 percent in 2019.[17] Such group proportions matter. To maximize the benefits and minimize the drawbacks of social diversity, demographic minorities should be included in sufficiently large numbers—a critical mass of around 30 percent—so that they do not fall prey to tokenism.

In the absence of dramatic changes in representation, the second solution is to make representative examples of the existing sample more available by increasing the visibility of small-N members and thereby changing what is salient. Role models—be they in the form of real-life leaders, speakers on panels and at events, characters represented in films and media, portraits on walls, or names of buildings, streets, and conference rooms—are powerful influences on behavior and help to counteract stereotype threat.[18] The "This Girl Can" campaigns in Australia and the United Kingdom, for example, provided realistic role models representative of the general population to encourage sport among women and girls.[19] In the absence of counter-stereotypical role models, other creative approaches can work. In computer science classrooms, replacing male-stereotyped Star Wars images with more gender-neutral nature landscapes on the walls has been shown to equalize female undergraduates' interest in computer science with that of their male counterparts.[20]

The third solution entails changes in decision-making processes. Joint evaluation whereby multiple candidates are assessed comparatively against each other, as opposed to individually in isolation, has been shown to reduce bias in decision-making.[21] Joint evaluation focuses evaluators' attention on individual-level information about each candidate and decreases their reliance on stereotypes. Similarly, simultaneous decisions whereby multiple candidates are selected at the same time rather than one at a time—whether in hiring, promotion, or election contexts—have been shown to lead to more diversity in outcomes.[22] Thus, instead of hiring for one open position in March, one in May, and one in October, companies would do better to hire for all three open positions at the same time, thereby benefiting from the ability to evaluate a larger pool of candidates comparatively and make three simultaneous hiring decisions.

17 John Beshears, Iris Bohnet, and Jenny Sanford, "Increasing Gender Diversity in the Boardroom: The United Kingdom in 2011 (A)," Harvard Business School Supplement 918-006, July 2019.

18 Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India," *Science*, 335, no. 6068 (February 2012): 582–586.

19 "This Girl Can," Sport England, accessed January 27, 2020, http://www.thisgirlcan.co.uk; "This Girl Can," Victorian Health Promotion Foundation, accessed January 27, 2020, http://www.thisgirlcan.au.

20 Sapna Cheryan et. al. "Ambient Belonging: How Stereotypical Cues Impact Gender Participation in Computer Science," *Journal of Personality and Social Psychology* 97, no. 6 (2009): 1045–1060.

21 Iris Bohnet, Alexandra van Geen, and Max Bazerman, "When Performance Trumps Gender Bias: Joint Versus Separate Evaluation," *Management Science* 62, no. 5 (2016): 1225–1234.

22 Edward H. Chang et. al., "The Isolated Choice Effect and Its Implications for Gender Diversity in Organizations," forthcoming in *Management Science* (n.d.).

References

Ambady, Nalini, Margaret Shih, Amy Kim, and Todd L. Pittinsky, "Stereotype Susceptibility in Children: Effects of Identity Activation on Quantitative Performance." *Psychological Science* 12, no. 5 (September 2001): 385–390.

Arrow, Kenneth J., "The Theory of Discrimination", in O. Ashenfelter and A. Rees (eds.), *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124, no. 4 (November 2009): 1497–1540.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India." *Science*, 335, no. 6068 (February 2012): 582–586.

Bertrand, Marianne, and Esther Duflo. "Review on Field Experiments on Discrimination." In *Handbook of Field Experiments*, edited by Abhijit Banerjee and Esther Duflo. North Holland, U.K.: Elsevier Science, 2017.

Beshears, John, Iris Bohnet, and Jenny Sanford. "Increasing Gender Diversity in the Boardroom: The United Kingdom in 2011" (A). Harvard Business School Supplement 918-006, July 2019.

Bohnet, Iris. *What Works: Gender Equality by Design*. Cambridge, MA: The Belknap Press of Harvard University Press, 2016.

Bohnet, Iris, Ashley Craig, and Clémentine van Effenterre, "Overcoming Gender Differences in Interview Ratings." Harvard Kennedy School Working Paper, Harvard University, Cambridge, MA, 2019. Microsoft Word file.

Bohnet, Iris, and Farzad Saidi. "Informational Inequity Aversion and Performance." *Journal of Economic Behavior and Organization* 159 (March 2019): 181–191.

Bohnet, Iris, Alexandra van Geen, and Max Bazerman. "When Performance Trumps Gender Bias: Joint Versus Separate Evaluation." *Management Science* 62, no. 5 (May 2016): 1225–1234.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. "Stereotypes." *The Quarterly Journal of Economics* 131, no. 4 (November 2016): 1753–1794.

Chang, Edward H., Erika L. Kirgios, Aneesh Rai, and Katherine L. Milkman. "The Isolated Choice Effect and Its Implications for Gender Diversity in Organizations." Forthcoming in *Management Science* (n.d.).

Cheryan, Sapna, Victoria C. Plaut, Paul G. Davies, and Claude M. Steele. "Ambient Belonging: How Stereotypical Cues Impact Gender Participation in Computer Science." *Journal of Personality and Social Psychology* 97, no. 6 (December 2009): 1045–1060.

Fryer, Roland, Devah Pager, and Jörg L. Spenkuch. "Racial Disparities in Job Finding and Offered Wages." *The Journal of Law & Economics* 56, no. 3 (August 2013): 633-689.

Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. J*udgment under Uncertainty*: Heuristics and Biases. Cambridge: Cambridge University Press, 1982.

Kanter, Rosabeth M. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *American Journal of Sociology* 82, no. 5 (1977): 965–990.

Norris, Pippa. "The Impact of Electoral Reform on Women's Representation." *Acta Politica,* 41, no. 2 (July 2006): 197–213.

Pager, Devah, and Diana Karafin. "Bayesian Bigot? Statistical Discrimination, Stereotypes, and Employer Decision Making." *The Annals of the American Academy of Political and Social Science*, 621, no. 1 (January 2009): 70–93.

Phelps, Edmund. (1972). "The Statistical Theory of Racism and Sexism." *American Economic Review* 62, no. 4 (1972): 659–61.

Phillips, Katherine W., and Phillips, Damon. "Nationality Heterogeneity, Performance, and Blau's Paradox: The Case of NHL Hockey Teams, 1988-1998." Paper presented at the Academy of Management conference, New Orleans, LA, August 2004.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. "Meta-analysis of Field Experiments Shows no Change in Racial Discrimination in Hiring over Time." P*roceedings of the National Academy of Sciences* 114, no. 41 (2017): 10870–10875.

Sport England. "This Girl Can." Accessed January 27, 2020. http://www.thisgirlcan.co.uk.

Victorian Health Promotion Foundation. "This Girl Can." Accessed January 27, 2020. https://thisgirlcan.com.au.