Discovering the Diversity-Innovation Paradox in the Sciences

Sociology of Science, Diversity Effects, and Research Innovation,
Stanford University

Introduction

This research was a collaborative effort between the Graduate School of Education, Department of Computer Science, and Department of Linguistics at Stanford University. Dutch researcher, Dr. B. Hofstra, Assistant Professor of Radboud University, the Netherlands worked alongside Daniel McFarland, Professor of Education and Sociology, Stanford University. The paper was supported by three NSF grants and one grant from the Dutch Organization for Scientific Research (NOW Grant).

The diversity paradox is the hypothesis that people from diverse backgrounds and perspectives can hasten innovation because "historically underrepresented groups often draw relations between ideas and concepts that have been traditionally missed or ignored," the researchers recently noted in their study "The Diversity-Innovation Paradox in Science."

The "paradox" is that although these groups contribute accelerated innovation in organizations, they also typically "have less successful careers within them."

With this in mind, the team from Stanford's Graduate School of Education, Department of Computer Science and Department of Linguistics wondered if the diversity paradox also applies in the sciences.

The question inspired a study focused on a dataset of 1,208,246 dissertations (collected within **ProQuest Dissertation & Theses Global**) filed by US doctorate-awarding universities from 1977 to 2015. The goal of the study was to understand "if gender and racially underrepresented scholars are likely to innovate and innovation supposedly leads to successful academic careers, then how do we explain persistent inequalities in scientific careers between minority and majority groups?"

The Project

In this day and age, researchers need rich longitudinal datasets that accurately reflect a population of subjects. With the digitalization, many research professors can now tap into datasets like these that weren't available 20-30 years ago. Professors, like Daniel McFarland, a professor of Education at Stanford University, see resources like ProQuest Dissertations & Theses Global as a nice example of such high quality data and use it in their research. McFarland was able to access the full text and metadata of the Dissertations & Theses by using TDM Studio to isolate the publication years and access the content for text and data mining.

McFarland studies the social and organizational dynamics of educational systems like universities and disciplines. McFarland is currently engaged on several different projects including writing a textbook on Social Network Analysis in R with Craig Rawlings, Jeff Smith, and James Moody. He also has generated a line of articles on the sociology of science, diversity effects, and research innovation. That's where using a product like PQDT Global has helped McFarland. Using PQDT Global and other resources, he's able to make his research as well-rounded as possible for publication.

According to McFarland, the standards for being published have risen so researchers have to check their models and inferences many different ways. McFarland also stated that "compared to 20 years ago, research standards for publication are higher. That's a good thing, but it also means we have to find rich longitudinal information that can be anchored in known populations – then our inferences make sense."

McFarland and his research team turned to PQDT Global because they wanted to study scholar careers and they wanted data that accurately reflected the pool of potential scholars. "What I liked about PQDT Global is that it wasn't biased like some journals are and it's evenly representative across disciplines. We don't get only journal science, or conference proceedings, or books, but rather see people of all fields who write theses. PQDT also doesn't select on the outcome. By this I mean most studies look at faculty or persons who got PhDs and succeeded in becoming faculty. With PQDT we have recourse to following all PhDs into various jobs and out of academia – so the entire pool of potential faculty." McFarland and his team needed a data set of at least 30 years and all potential future faculty to develop models of "successful" academic careers. This is something that PQDT can offer with content that dates back to the 1700s.

"ProQuest Dissertations & Theses Global is a wonderful corpus for scholars to work with, and especially if they are interested in following the pool of potential researchers and their careers," McFarland stated.

"What I like about PQDT is that it affords us a sizeable sample of beginning researcher's careers. Using National Center of Education Statistics census of all US PhDs, we can weight the PQDT sample and make inferences about all US PhDs."

Results

Using text analysis and machine learning (via ProQuest **TDM Studio**), the research team explored three key areas:

- How do we detect scientific innovations?
- Are underrepresented groups more likely to generate scientific innovations?
- Are the innovations of underrepresented groups adopted and rewarded?

"Our analyses show that underrepresented groups produce higher rates of scientific novelty," the Stanford team concluded. "However, their novel contributions are devalued and discounted."

"For example," researchers explained:

Novel contributions by gender and racial minorities are taken up by other scholars at lower rates than novel contributions by gender and racial majorities, and equally impactful contributions of gender and racial minorities are less likely to result in successful scientific careers than for majority groups.

"These results suggest there may be unwarranted reproduction of stratification in academic careers that discounts diversity's role in innovation and partly explains the underrepresentation of some groups in academia," according to their published paper.

Read the paper in its entirety, published in 2020 by PNAS (Proceedings of the National Academy of Sciences of the United States of America), or learn about how researchers at Stanford and other universities are mining ProQuest dissertations using the TDM Studio analysis tools to uncover patterns and trends in advanced scholarship.

About TDM Studio

ProQuest's workflow solution for text and data mining is designed for research, teaching and learning. TDM Studio provides access to sought-after content including current and historical newspapers, primary sources, scholarly journals, and dissertations and theses. It empowers researchers, students and faculty to analyze documents by uncovering connections and patterns that lead to career-defining discoveries.

References

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The Diversity–Innovation Paradox in science. *Proceedings of the National Academy of Sciences, 117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Learn more at www.proquest.com/go/tdm-studio
or contact your ProQuest representative today.

about.proquest.com

To talk to the sales department, contact us at
1-800-779-0137 or **sales@proquest.com**.

ProQuest®

Part of **Clarivate**