

NBER WORKING PAPER SERIES

MARGINALIZED AND OVERLOOKED? MINORITIZED GROUPS AND THE ADOPTION
OF NEW SCIENTIFIC IDEAS

Wei Cheng
Bruce A. Weinberg

Working Paper 29179
<http://www.nber.org/papers/w29179>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2021, Revised April 2023

Cheng acknowledges funding support from the National Natural Science Foundation of China (NSFC) grant 71803047 and Shanghai Pujiang Program through grant 2019PJC022. Weinberg is grateful for support from NSF EHR DGE grants 1760544 and 1535399, from R01 GM140281, and NIA, OBSSR and NSF SciSIP through P01 AG039347. All projects supported data construction and analysis. The research aligns most closely with the substance of 1760544 and R01 GM140281. Weinberg was supported on P01 AG039347 by the NBER directly and on a subaward from NBER to Ohio State. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Wei Cheng and Bruce A. Weinberg. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Marginalized and Overlooked? Minoritized Groups and the Adoption of New Scientific Ideas
Wei Cheng and Bruce A. Weinberg
NBER Working Paper No. 29179
August 2021, Revised April 2023
JEL No. D85,J71,O31

ABSTRACT

The diffusion and use of new ideas is critical for producing innovations and realizing their potential. We explore how characteristics of innovators and potential adopters affect the adoption of important, new scientific ideas in networks. Using rich data on biomedical researchers and their networks, natural language processing, and a novel two-way fixed effects strategy, we find that new ideas introduced by female scientists are under-utilized for two reasons. First, female innovators are less connected than men. Second, at short network distances, researchers (especially men) adopt women's ideas less. Similar gaps hold for underrepresented racial and ethnic minorities.

Wei Cheng
School of Business
East China University of Science and Technology
130 Meilong Rd, XuHui
Shanghai, China 200037
chengwei@ecust.edu.cn

Bruce A. Weinberg
The Ohio State University
Department of Economics
410 Arps Hall
1945 North High Street
Columbus, OH 43210
and NBER
weinberg.27@osu.edu

1. Introduction

Do women and underrepresented racial and ethnic minorities receive less attention than men and majority groups for their contributions (Wenneras and Wold, 1997; Halpern et al., 2007; Koffi, 2021)? This paper explores this question in an interesting and consequential population - biomedical scientists. Scientists are ideal for studying idea flows because they are highly knowledgeable and strongly incentivized to keep abreast of and adopt the latest advances. Moreover, the contents of scientists' contributions are publicly available, their contributions can be identified from authorship, and their networks can be reconstructed. With the population of scientists becoming increasingly diverse, disparities in the adoption of ideas are becoming increasingly costly for the scientific enterprise. Biomedical scientists are particularly consequential because research on conditions relevant for underrepresented groups is frequently undertaken by members of those groups, so a failure of research by members of these groups to receive appropriate attention would likely increase underlying health disparities (Oh, et al., 2015; Association of American Medical Colleges, 2009).

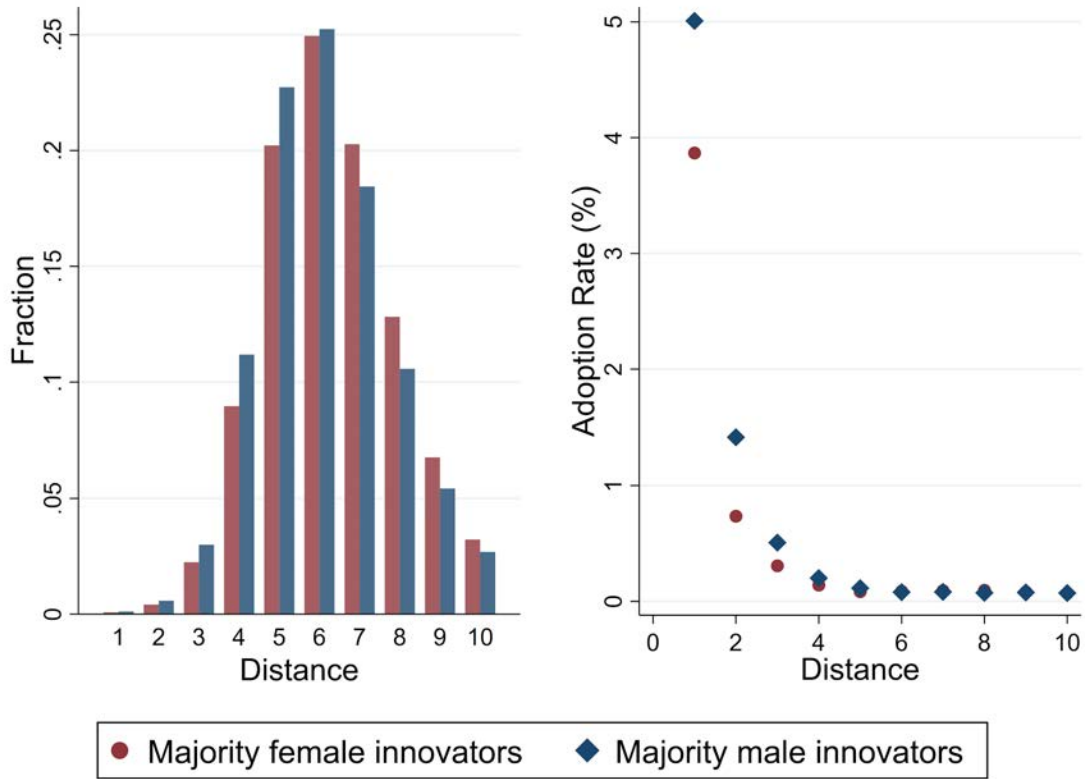
This paper combines two methods to explore differences by gender, race, and ethnicity in the adoption of ideas. First, we employ recent methods in natural language processing to identify the most important ideas (one-, two-, and three-word combinations, “n-grams”) introduced each year between 1980 and 2008 and track their utilization at population scale for nearly a quarter of a century (see Packalen and Bhattacharya, 2019; Staudt et al., 2018). We do so using the text of titles and abstracts in MEDLINE, a massive database of biomedical research articles with nearly 30 million articles dating back to the 19th century. The ideas we study are important. Among many other important advances, they include the work on HIV / AIDS that has transformed it from an

incurable condition to a chronic, manageable disease. They also cover the development of a wide range of methods and findings in molecular biology that are the foundation for the biotech revolution. Second, we build on the well-known idea that technology adoption declines with distance. Using these building blocks, we identify as “innovators” all the authors of the article(s) that first introduce an idea that becomes important and a set of “potential adopters”. We then construct the distance of each potential adopter to the innovators of each idea and estimate how adoption of ideas decays with network distance based on the characteristics of innovators and potential adopters.

Figure 1 illustrates our approach in the case of gender, with implications for women’s health. The left panel shows that researchers are on average further from ideas generated by teams with more women than men (red bars) compared to ideas generated by teams with more men than women (blue bars). This finding implies that female innovators are not as well-connected as male innovators.⁴ What further exacerbates disparities in adoption is that at each distance, ideas introduced by female majority teams are less likely to be adopted (right panel). Even at a distance use of 1, people are more likely to choose to build on ideas introduced by majority male teams. Both effects lead to lower adoption rates for female majority new ideas. Specifically, the 5-year adoption rate of new ideas from male majority teams is 25% higher than that of ideas from female majority teams (even among ideas in the top 0.01% of ideas).

⁴ Throughout this paper, we use female and male for gender adjectives and woman and man as gender nouns, but our intention is not to distinguish biological sex from gender identity. We further note that our gender concept is static and binary, which is a limitation of our work. Unfortunately, we are not aware of gender imputations that account for non-binary or fluid genders.

Figure 1. Gender composition of innovator team, network position, and adoption



Notes: All new ideas are divided into two groups: ideas introduced by teams with a majority of female innovators (red) and ideas introduced by teams with a majority of male innovators (blue). The left panel plots the distribution of researchers' network distance to the two groups of ideas separately. The right panel shows the 5-year adoption rates for ideas from both sets of teams in percentage points at each network distance.

Although we have no reason to believe it is the case, some readers may wonder if our finding that the ideas introduced by women are adopted less than those introduced by men arises because the ideas introduced by women are inherently different (e.g. not as promising or widely applicable) from those introduced by men. To address this possibility, we rely on two key features of our context: (1) that articles typically have more than one author and (2) that network proximity is a strong determinant of knowledge flows. In our estimation, for each idea and each potential adopter, we identify the nearest individual innovator. We estimate regressions that control for idea fixed effects that eliminate all differences across ideas in the applicability of the idea to potential

adopters. Intuitively, each active biomedical researcher (i.e. potential adopter) will be closer to one (or more) of the innovators than to others. Thus, we can leverage the network structure, measured prior to the introduction of the idea, to identify whether, for each idea and potential adopter, the “closest” innovator to that potential adopter is a woman or a man, and see whether, controlling for the distance to the closest innovator, a potential adopter is more likely to adopt an idea when the closest innovator is a woman or a man. In this way, we use the presence of multiple innovators and network proximity to control for the “quality” of ideas.

Another potential explanation for our finding that the people who are closer to women innovators are less likely to adopt ideas is that the people closer to women innovators are inherently less likely to adopt new ideas. For instance, women tend to be closer to other women, and women may be less likely to adopt new ideas, perhaps because they occupy positions that are less central in networks (and hence have less access to complementary innovations) or because of less access to resources (Ginther and Kahn, 2009).⁵ We address these concerns by controlling for the direct effect of network distance. We also observe adoption behavior for multiple ideas for potential adopters and include fixed effects that account for differences in the adoption propensity across potential adopters. In this way, we control for unobserved differences in adoption propensities for each potential adopter. Our estimates indicate that, at any given distance, women and underrepresented minorities (URMs) are less likely to adopt new ideas. However, controlling for these differences and potential adopter fixed effects accounts for only a minority of the disparity in the adoption of ideas of women innovators.

⁵ Another potential concern is that our sample of ideas is itself determined by adoption rates, so that if the characteristics of innovators do indeed determine adoption, the characteristics of innovators may well affect inclusion in our sample. We acknowledge that such a process operates, but note that we (1) control for idea fixed effects, so that our estimates are identified from within-idea variations and (2) that this process implies that the ideas from, for instance, predominantly women innovator teams in our sample are, if anything, “better” ideas than those from predominantly men innovator teams.

Our work contributes to the literature in three ways. First, we contribute to the literature on underrepresentation in science. Gender and racial disparities have long been noted in history and sociology of science, and they are receiving increasing interests among economists. A sizeable body of work suggests that women and under-represented minorities (URM's) in various areas of science fall behind in many dimensions. Women experience greater barriers as their careers progress. They are found to be disproportionately less likely to complete M.D. and Ph.D. degrees (Baker, 1998); receive tenure and be promoted to full professor (Ginther and Kahn, 2009); or be awarded professional awards and prizes (Lincoln et al., 2012). Moreover, they seem to perform worse in terms of publication and patent rates (Ding et al., 2006), ⁶ obtaining research funding (Ley and Hamilton, 2008; Ginther et al., 2011), and earnings (Buffington et al., 2016). As scientific inquiry is becoming increasingly collaborative over time (Wuchty et al., 2007), women are found to engage less in collaborative research, especially international collaborations (Abramo et al., 2013). Using a global and cross-disciplinary bibliometric analysis, Larivière et al. (2013) finds that male dominance in research output and first authorship exists in most countries in the world.

At the same time, women are often under-recognized for their scientific contributions (Rossiter, 1993). Recently, researchers have focused on the credit that women receive for their contributions in the form of author order (e.g. Cikara et al., 2012, West et al., 2013, Lerchenmüller, Lerchenmüller, and Sorenson, 2018, Marschke, et al., 2018), in terms of inclusion as an author at all (Ross et al. 2021), or credit for authorship (e.g., Sarsons et al., 2021) and citations (Ghiasi et al., 2016; Beaudry and Larivière, 2016; Hofstra et al., 2020; Koffi, 2021). We enrich this line of study by exploring the possibility that the ideas produced by women and underrepresented groups

⁶ However, Card et al., (2020) finds that gender biases are negligible in referees' and editors' evaluation of female-versus male-authored papers in economics.

are less likely to be adopted than those of men and members of majority groups, which may help explain differences in citations, although citation patterns known to be skewed by artifacts of the citation process (discussed more below, including in footnotes 7 and 8).

Beyond the harm experienced by individual researchers, gender and racial disparities in science can generate substantial loss to society by slowing scientific progress, especially for conditions affecting underrepresented groups. Moreover, adverse experiences may deter entry or hasten exit of members of underrepresented groups at a time when society is seeking to increase growth and welfare through innovation, and diversity itself is often associated with higher research quality (Reagans and Zuckerman, 2001; Freeman and Huang, 2014).

Our second contribution connects to the literature on knowledge spillovers. Technology and knowledge spillovers have been central to economists' understanding of growth, trade, and urban agglomerations for over thirty years (Romer, 1986; Lucas, 1988; Krugman, 1991; and Glaeser, et al., 1992). Unfortunately, the micro evidence for such spillovers tends to be indirect (e.g. inferred from productivity) rather than measured directly and is surprisingly weak given the critical importance of knowledge spillovers (Azoulay, Graff-Zivin and Wang, 2010; Waldinger, 2010, 2012; Borjas and Doran, 2012, 2015; and Catalini, 2017).⁷ One novel aspect of our work relative to this literature is that we directly measure knowledge flows using text analysis. At least since the seminal work by Jaffe, Trajtenberg and Henderson (1993), researchers have sought to infer knowledge flows using citations, frequently in patent records across countries / regions (Peri, 2005), between firms (Almeida and Kogut, 1999; Gomes-Casseres, 2006), and among scientists (Singh, 2005). Text analysis has a number of advantages over this approach - it allows us to track

⁷ Kaiser (2005) is a fascinating exception, showing how the use of Feynman diagrams diffused through the physics community from the different ways in which physicists drew them.

the use of specific ideas and avoid artifacts in the citation process in both patents⁸ and the scientific literature.⁹

Lastly, our analysis of coauthorship networks allows us to focus on the transmission of ideas across individuals who have connections with one another. Given that people's knowledge of new ideas is imperfect, researchers close to the innovators of a new idea benefit from the ability to obtain tacit knowledge at a lower cost. The literature offers abundant evidence regarding the importance of *spatial proximity* in knowledge flows (e.g. Jaffe, et al., 1993; Almeida and Kogut, 1999; Thompson and Fox-Kean, 2005, Waldinger, 2010, 2012; Borjas and Doran, 2012, 2015; Catalini, 2017; Ham and Weinberg, 2021). However, most of the mechanisms underlying the role of spatial proximity, including a higher chance of meeting, lower communication costs, and enhanced trust, hold for *network proximity* as well. Despite its potential as a mediator, the literature on network proximity for idea diffusion is relatively limited (Singh, 2005; Zacchia, 2019) and our two-way fixed effects design has the potential for facilitating future network analyses.

2. Data

2.1 Data

Our primary data source is MEDLINE, the U.S. National Library of Medline's (NLM) premier bibliographic database, which is akin to Econlit for biomedical research. MEDLINE contains over 30 million biomedical research articles from around the world dating back to the 19th century, with

⁸ Alcácer and Gittelman (2006) show that more than half of the citations on a patent are added by patent examiners, and thus may not represent knowledge flows.

⁹ Following Merton (1973), scientific citations are taken to reflect knowledge flows, but there is considerable evidence for a variety of other motives for citations (see Gilbert, 1977; Moed, 2005; Moed and Garfield, 2004; Collins, 2004; Woolgar, 1991; Fong and White, 2017; Koffi, 2021). Moreover, different types of articles receive different numbers of citations (Moed and van Leeuwen, 1995).

coverage that improves greatly over time, especially after 1966. In addition to the comprehensiveness of the MEDLINE data, another strength is that it provides rich information on journal articles, including authors' names, publication year, and the text of titles and (increasingly since 1975) abstracts, which we use in our text analysis. MEDLINE also includes Medical Subject Headings (MeSH) for each publication that indicate its main research areas. We use these to identify potential adopters of a new idea by identifying researchers who have conducted research on similar topics as the innovators who first introduced the idea.

Another crucial data source is the “Author-ity” MEDLINE author disambiguation database, updated through 2018 (Torvik, Weeber, Swanson, and Smalheiser, 2006; Torvik and Smalheiser, 2009). In a corpus the size of MEDLINE, identifying authors solely based on names is likely to generate lumping errors, where people with similar names are treated as one person, and splitting errors, where one author is regarded as different people when names change or are written in different ways. “Author-ity” provides a high-quality, algorithmic disambiguation of MEDLINE that allows us to identify each scientist's publication records with a high level of accuracy.

In this paper, we focus on U.S.-based scientists. To identify author locations, we use MapAffil (Torvik, 2015), which provides affiliation information for MEDLINE authors. We eliminate people who are ever affiliated with organizations outside the U.S. Additionally, we supplement our data with Genni (Torvik and Agarwal, 2016), which predicts gender, and Ethnicolr (Laohaprapanon and Sood, 2017), which predicts race and ethnicity based on scientists' names. Genni provides ethnicity-specific gender predictions. Ethnicolr classifies people into four categories that combine race and ethnicity: Hispanics (of any race), non-Hispanic Asians, non-Hispanic Blacks, and non-Hispanic Whites.

2.2 Important of new ideas

Following Packalen and Bhattacharya (2019) and Staudt et al. (2018), we extract the main concepts or ideas contained in a paper by taking the words (unigrams), two-word phrases (bigrams) and word triplets (trigrams) in the text of the title and abstract. A new idea is an n-gram that first appeared in a certain year and that had never been used before. We refer to this year of first appearance as the birth year of a new idea. We focus on new ideas born between 1980 and 2008. This time window is chosen for two reasons. First, we exclude ideas that are introduced before 1980 to allow for a sufficient “burn-in” period in the calculation of an idea’s birth year. Many old terms will be identified as having a birth year when the data start¹⁰. However, after over ten years, a first-appearing term is likely to be an innovation. Second, the last idea birth year we choose is 2008 so that we have at least 10 years (up to 2018) after an idea is introduced to measure its adoption.

One potential problem with this approach of using new terms to represent new ideas is that some new terms are merely new combinations of words but do not constitute important innovations. To avoid such cases, which are rarely widely adopted, we rank the ideas by the number of total mentions within each birth year and take the top 0.01% of the distribution for each idea birth year. Therefore, the final list of new ideas we focus on is a set of the most important new ideas in biomedicine introduced by U.S. scientists over nearly 3 decades. For example, ideas ranked highly include well-known breakthroughs and advances in biomedical research such as polymerase chain

¹⁰ In the early years of the data, terms such as “blood” have very high frequencies of mentions. However, they are not innovations but commonly used language in biomedicine. Hence, it is important to have a sufficiently long initial period to set them apart from new ideas.

reaction, small interfering RNAs, and HIV / AIDS. The number of the important new ideas in our list ranges from 63 in 1980 to 188 in 2008.

2.3 Sample construction

For each new idea (and its innovators), we identify potential adopters by selecting all other researchers who have works with at least one MeSH term in common with the innovators before the birth of the new idea. Formally, to construct our sample for idea i that first appears (i.e. is born) in year τ , we let $I_{i,\tau}$ denote the set of innovators who originated idea i in year τ . Let $M_{i,\tau}$ represent all MeSH terms that $I_{i,\tau}$ have published on in or before year τ . Our set of potential adopters for idea i , $A_{i,\tau}$, is defined as the set of researchers who have published on at least one MeSH term in $M_{i,\tau}$ in or before year τ and who are not in $I_{i,\tau}$ itself.

Next, for each potential adopter in $A_{i,\tau}$, we calculate his/her shortest path to any of the innovators of idea i , $I_{i,\tau}$, based on the collaboration networks in the year the idea was introduced and the two preceding years, $\tau - 2$ to year τ ¹¹. (Adoption is tracked starting the year after τ , i.e. from $\tau + 1$.) Network distance represents how close a potential adopter is to a new idea in collaboration networks. (We truncate the network distance at 10 because over 90% of all finite distance pairs have a distance equal to or less than 10 and adoption rates at large distances are extremely low.) Two time frames are used to measure adoption: 5 years (i.e. starting from $\tau + 1$ through $\tau+5$) and 10 years (i.e. starting from $\tau + 1$ through $\tau+10$) after the introduction of a new idea. The primary results presented in the paper are based on adoption within 5 years because the effects of pre-existing network proximity likely decline as time passes. Analyses of adoption within 10 years are

¹¹ Since network links are observable only when publications are generated, and it is common for a scientist to have zero publications in some years, we choose a three-year window around the birth of a new idea to define the networks. Infinite network distance is not considered in this analysis.

conducted as robustness checks. More specifically, the adoption outcome variable represents independent adoption defined as using a new idea in an article that does not have any of the original innovators as a coauthor because “adoption” through coauthorship with one of the innovators of the idea likely operates differently from independent adoption.¹²

Our unit of analysis is an idea-adopter pair, and our sample construction procedure yields over 76 million observations. We take a 10% random sample as our analytical dataset. Our final analytical sample, which is described in Table 1, consists of 7,600,166 observations, corresponding to 3,427 new idea terms and 536,813 unique potential adopters. Of all potential adopters, 52% are men, 33% are women, and the rest, 15%, have uncertain gender predictions. In terms of race and ethnicity, the largest group is non-Hispanic White (82.8%), followed by non-Hispanic Asian (12.6%), Hispanic (4.0%), and non-Hispanic Black (0.6%). These demographic features do not directly measure the overall demographics of the biomedical research workforce in that our sample only includes people who are at a distance less than or equal to 10 from the innovators of the new important ideas and weights people who are in areas with more important new ideas more heavily than others.

3. Results

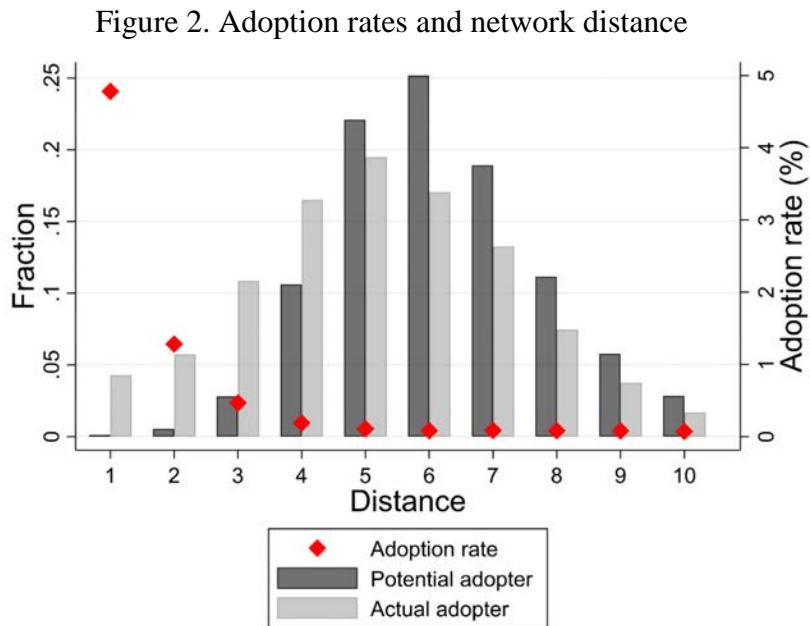
We present our results in three steps. First, we provide some graphical evidence of different patterns of idea adoption for researchers from various demographic groups. Second, we conduct a formal regression analysis and discuss potential mechanisms for our findings. Lastly, we perform

¹² The results are essentially identical using adoption or independent adoption as the outcome variable.

a series of decompositions to assess the relative contributions of network distance, the gender mix of adopters, and adoption conditional on distance, to the gender gap in idea adoption.

3.1 Descriptive results on gender

Figure 2 illustrates the relationship between network distance and the adoption of ideas. The red squares represent the probability of adoption within 5 years of introduction (right axis) at various distances from the people who introduced the new ideas (innovators). Not surprisingly (and surely not entirely for causal reasons), the adoption rate of important new terms declines sharply with network distance¹³. The bars and the left axis in Figure 2 show the distribution of network distance for all potential adopters (black bars) and actual adopters (grey bars). The grey bars are shifted left-ward, indicating that actual adopters are closer to innovators in network terms because adoption rates are highest at low distances.



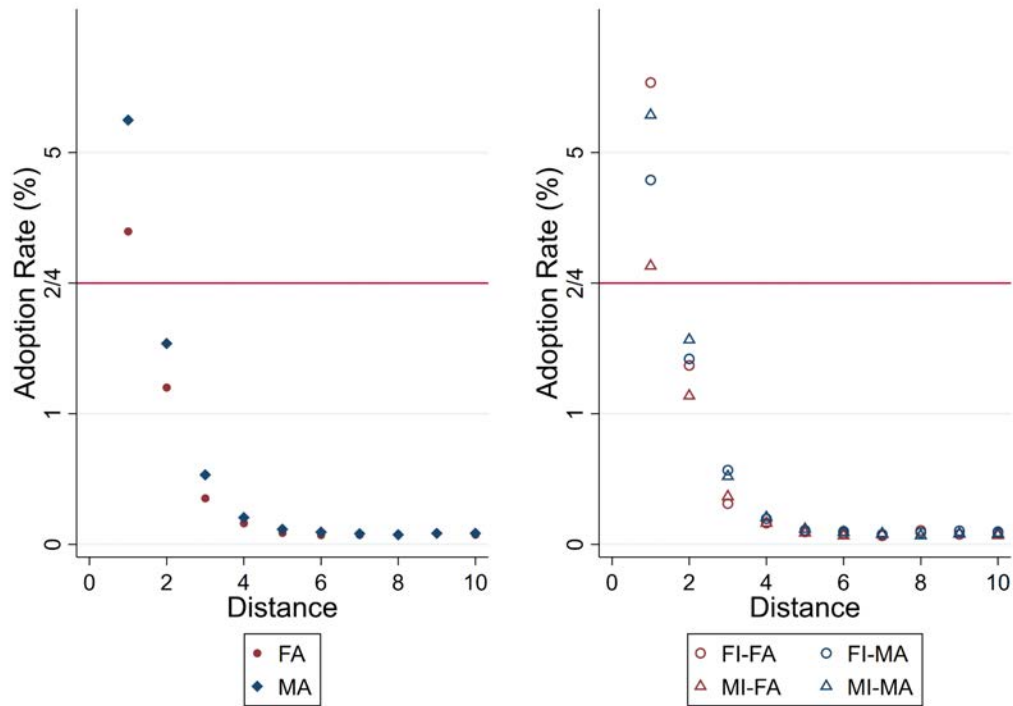
Notes: The red diamonds (corresponding to the right y-axis) represent the probability of adoption within 5 years following an idea's introduction by potential adopters at various network distances from the nearest innovator of an

¹³ It is noteworthy that while the adoption rate falls with distance, the adoption frequency is greatest at moderate distances because the number of potential adopters at moderate distances are orders of magnitudes higher than those at distance 1 or 2 from innovators.

idea. The bars show the distribution of a potential adopter's network distance to the nearest innovator of each important new idea. The black bars represent all potential adopters and the grey bars represent actual adopters.

One reason women's ideas may be less likely to be adopted than men's ideas is that women's ideas may be more likely to be adopted by other women and women may be less well integrated into networks. In other words, given that ideas are disproportionately adopted by those at short network distances, if the adopters of the ideas introduced by women are disproportionately women and, if women adopters are less well connected in networks, the ideas of women innovators may be further from the potential adopters with the greatest adoption propensity. We explore this possibility. Our data shows that women and men are at similar network distances from new ideas. The mean distance to new ideas is 6.15 for women and 6.12 for men (both have a median distance of 6). However, the left panel of Figure 3 (which has a spliced Y-axis for readability) shows that women are less likely to adopt important new ideas, especially at the shortest distances.

Figure 3. Gender and adoption rates



Notes: The left panel shows the adoption rates for female and male potential adopters at each network distance. The right panel considers adoption rates by different gender combinations of the potential adopter and nearest innovator. *FA* / *MA* indicate a female / male potential adopter. *FI* / *MI* indicate that the closest innovator is female / male respectively. We move the points at distance 1 down by two units for readability.

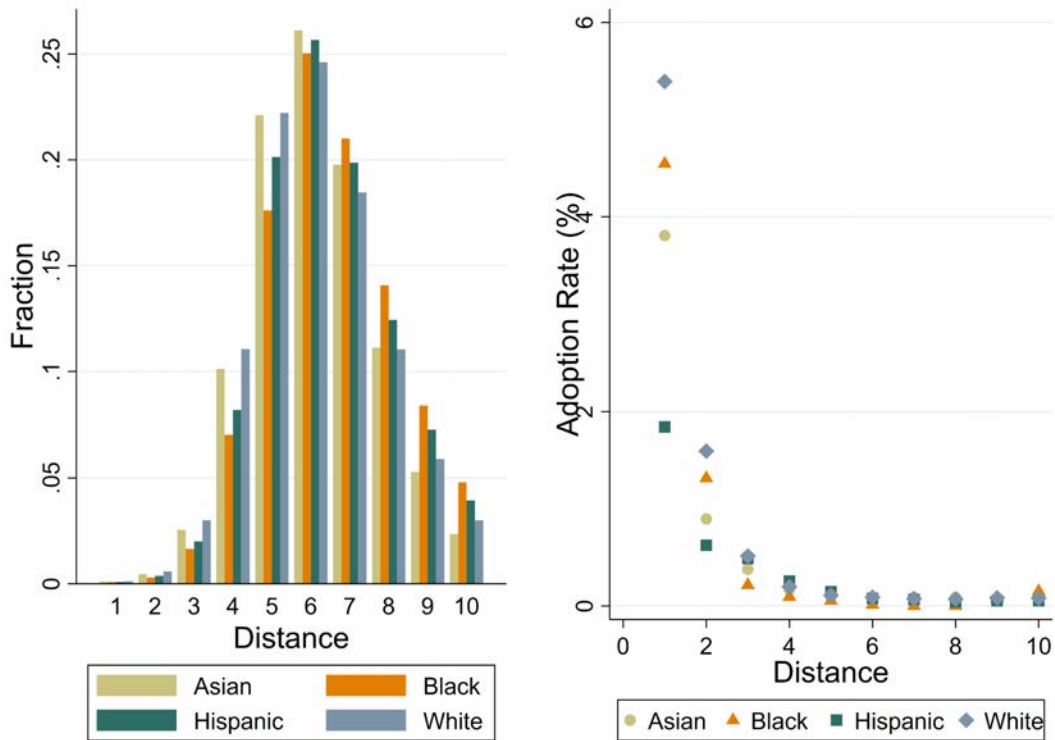
Given that women are less likely to adopt new ideas, the gender composition of potential adopters at each distance to innovators is potentially an important determinant of adoption. Let *FI* and *MI* represent whether, for a given potential adopter, the closest innovator of a given idea is female or male. *FA* and *MA* indicate the gender of a potential adopter, yielding four combinations: *FI-FA*, *FI-MA*, *MI-FA* and *MI-MA*. The right panel of Figure 3 (which has a spliced Y-axis for readability), depicts the adoption rates for each of the four combinations by network distance. Two points are noteworthy. First, for both genders, people are more likely to adopt the ideas of network-proximate innovators of their own gender. Second, while at longer distances men are more likely to adopt the ideas of male innovators than women are to adopt the ideas of female innovators, at short distances (1 and 2), women are slightly more likely to adopt the ideas of a female innovator than men are to adopt the ideas of a male innovator. Thus, it is not the case that women are uniformly less likely to adopt new ideas of women than men are to adopt the ideas of men. Interestingly, at short distances, the lowest adoption rate is by women potential adopters of the ideas of male innovators, followed by men's adoption of women's ideas. Overall, men's lower adoption of women's ideas – and there are more male than female biomedical authors – explains the lower adoption of women's ideas, and women's reluctance to adopt men's ideas contributes to women's lower adoption of new, important ideas.

3.2 Descriptive results on race/ethnicity and experience

While our primary analysis is by gender, our methods also apply to race and ethnicity. Figure 4 provides estimates by race and ethnicity. Given that Asians, Blacks, and Hispanics are underrepresented, we classify an idea as having a substantial minority (Asian/Black/Hispanic)

component if at least a fifth of innovators belong to that group. (The average idea is introduced by 5 innovators, so for the average idea, we are effectively assuming that 1 of the innovators is a minority.) Ideas that do not make that threshold are classified as non-Hispanic, White. The left panel of Figure 4 shows the distribution of potential adopters' network distance for the four groups of ideas. Ideas with Asian and White innovators are similar in terms of network distance; however, ideas with at least 20% Hispanic or Black innovators are farther away from potential adopters. In terms of adoption rates, the right panel shows that, at distances 1 and 2, the ideas of White inventor teams have the highest adoption rates, followed by ideas from Black, Asian, and then Hispanic innovators.

Figure 4. Race/ethnicity of the innovator



Notes: The left panel shows the distribution of network distance of potential adopters to the innovators of ideas by race and ethnicity of the innovators. If at least 20% of the innovators of an idea are Hispanic, we categorize the ideas as having a substantial Hispanic contribution. If less than 20% of the innovators are Hispanic, but 20% or more of the innovators are Black, we categorize this idea as having a substantial Black contribution. If neither of these conditions

is met, and at least 20% of the innovators are Asian, then the idea is identified as having a substantial Asian contribution. The remaining ideas are considered non-Hispanic White. The right panel shows the adoption rate of ideas for these four groups by potential adopters at each network distance.

3.3 Regression results

We investigate how innovator characteristics, potential adopter characteristics, and network distance interact to determine idea adoption rates using regressions of the form:

$$\begin{aligned}
 A_{i,j} = & \alpha_0 + \sum_{d=1}^5 \alpha_d \cdot 1(\text{Distance}_{i,j} = d) + \sum_{d=1}^5 \beta_d \cdot 1(\text{Distance}_{i,j} = d) \cdot X_j^{PA} \\
 & + \sum_{d=1}^5 \eta_d \cdot 1(\text{Distance}_{i,j} = d) \cdot X_{i,j}^I + \sum_{d=1}^5 \lambda_d \cdot 1(\text{Distance}_{i,j} = d) \cdot X_j^{PA} \cdot X_{i,j}^I + \gamma_i + \rho_j + \\
 & u_{i,j}
 \end{aligned} \tag{1}$$

where $A_{i,j}$ represents whether potential adopter j independently adopts idea i within five (or ten) years of its first introduction, and $\text{Distance}_{i,j}$ represents researcher j 's network distance to the closest innovator on idea i in the three year time window before its birth year. We include a full set of distance dummies to account for a flexible pattern in the relationship between network distance and adoption rates. Distances greater than 5 have similar coefficients, and have been combined into one group, which is the excluded / reference group. The term X_j^{PA} represents potential adopter j 's characteristics such as gender and experience. The term $X_{i,j}^I$ represents the characteristics of the innovator on idea i who is closest to potential adopter j .¹⁴ As discussed, our models contain idea fixed effects, represented by γ_i , which account for, among other factors, idea birth year differences, and potential adopter fixed effects, represented by ρ_j .

¹⁴ Averages are taken if there are multiple innovators who are the same distance from a given potential adopter.

There are four parameters of interest. First, the α_d capture how adoption declines with network distance. Given the inclusion of idea fixed effects, the α_d are identified from differences across people who are at different initial distances from the innovators of ideas. Ultimately, the α_d likely combine the causal effect of distance with the decline in the relevance of work as proximity to the innovators of an idea increases. Second, the β_d indicate how the adoption of ideas declines with distance from the nearest innovator according to the characteristics of the potential adopter. We also caution against a causal interpretation of these estimates given the potential for unobserved differences across potential adopters. Third, the η_d represent how the characteristics of the closest innovator interact with network distance. Given the inclusion of distance and idea fixed effects, the η_d are identified from how the adoption probability for the same idea changes for potential adopters who are closest to a male innovator compared to those who are closest to a female innovator as distance increases. Lastly, the λ_d describe how the interaction between innovator characteristics and potential adopter characteristics vary with network distance. Given the inclusion of potential adopter fixed effects (as well as distance and idea fixed effects), if X represents gender, the λ_d are identified from the decline with distance in the probability of adopting an idea for a male potential adopter compared to a female potential adopter when the closest innovator is a woman versus a man controlling for the propensity of each potential adopter to adopt ideas and for the adoption of the idea.

A unique feature of our regression analysis is that we take advantage of the team aspect of innovation to account for unobserved differences across ideas (e.g., in the quality or generality of the idea) captured by γ_i . For a potential adopter, by focusing on the closest innovator, we have variation in innovator characteristics even after conditioning on idea fixed effects. Similarly, we control for unobserved differences across adopters in adoption propensities through adopter fixed

effects (i.e., ρ_j). Thus, our identification leverages within-idea differences across potential adopters connected to the idea through different innovators and within-adopter differences across ideas.

Table 2 provides estimates for our baseline regressions without interacted distance variables. We gradually expand the specification to incorporate additional control variables. All estimates are multiplied by 100, so the coefficients can be interpreted as percentage point changes. There are three main findings. First, the adoption probability decreases monotonically and substantially with network distance. The large magnitude of the coefficient for distance 1 indicates that immediate neighbors of the innovators of a new idea are substantially more likely to adopt an idea than those who are farther away. Including both sets of fixed effects does little to the coefficients on the distance dummies.

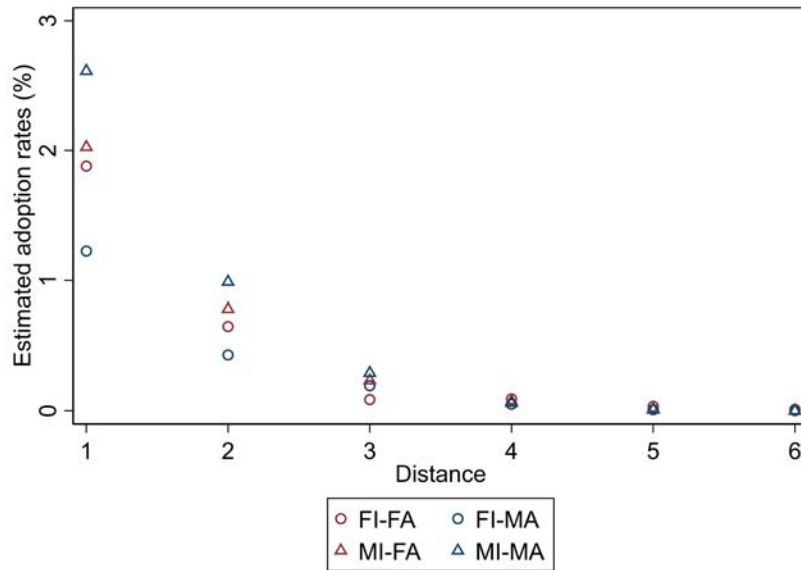
Second, potential adopters closest to an innovator who is female, Asian, or Hispanic, are less likely to adopt that idea. But the estimates become insignificant after we control for both individual fixed effects and idea fixed effects. Lastly, in terms of characteristics of potential adopters, female scientists are 0.029% less likely to adopt new ideas in the first 5 years. Considering that the overall adoption rate is 0.1%, this gender effect is quite sizable, although it may well reflect other differences that are related to gender. Compared to the reference group of Whites, Blacks and Hispanics are less likely to adopt new ideas. (These effects are absorbed in the individual fixed effects when they are included in Columns 2 and 3.)

Table 3 shows how the effect of network distance varies by the gender of the innovator and potential adopter. The distance coefficients are qualitatively similar to those in Table 2. The negative coefficients on the interaction terms between network distance and the closest innovator

being a woman suggest that being close to a female innovator significantly decreases a researcher's probability of adopting a new idea. Similarly, the interactions between network distance and the potential adopter being a woman indicate that even when women are proximate to innovators of a new idea, they are less likely to adopt the idea, which is consistent with Figure 3.

Interestingly, the model with three-way interactions between network distance, closest innovator gender, and potential adopter gender (reported in Column (3) of Table 3 and plotted in Figure 5) shows that, holding all else constant and compared to men, women are about as likely to adopt new ideas from female innovators as men are to adopt ideas from male innovators at short distances (1 and 2). However, even at short distances, men are the least likely to adopt women's ideas among all four gender configurations. As network distance increases, the disparity in adoption rates for different gender combinations becomes insignificant. In other words, women's ideas are even more restricted to nearby female potential adopters. But women are still a minority in the biomedical research workforce. These findings suggest that low adoption rates for those closest to women innovators arise because men are less likely to adopt women's ideas even though they are only a step or two away from the innovators.

Figure 5. Regression Estimates of Adoption by Gender of Innovator and Adopter



Notes: The figure plots regression-based adoption rates from Column 2 of Table 3 for different combinations of the gender of a potential adopter and his or her closest innovator. *FA* / *MA* indicate a female / male potential adopter. *FI* / *MI* indicate that the closest innovator is female / male respectively.

One may be concerned that the observed gap in adoption rates for men's and women's ideas is due to differences in the "quality" of new ideas introduced by men versus women. Two points are worth noting: first, our list of new ideas is from the very top of all new ideas - they are all of high quality. Second, our regression specification is at the idea-potential adopter pair level; and we incorporate idea fixed effects, which absorb differences in the quality of different ideas. Nevertheless, there may exist some unobserved idea characteristics that correlate with both innovator gender and network distances. One natural possibility would be the prestige of the journal in which the idea is originally published, which may reflect the quality of the idea or its visibility to potential adopters. To address this possibility, we construct a measure of the ranking of the journal that first published the new idea. More specifically, we use data from Scimago Journal Rank, which represents the average number of weighted citations received per document

published in a journal during the previous three years¹⁵. Then we include the interaction of this journal rank and network distances in our regressions. The estimates are reported in Column (3) of Table 3, and they are essentially unchanged. Therefore, our results are not driven by confounding factors associated with the quality or visibility of new ideas.

To probe the robustness of our results to the 5-year window for adoption, we repeat our estimation using a binary variable for adoption within 10 years after the birth of a new idea as the outcome variable. The estimates are reported in Columns (4)-(6) of Table 3. The pattern of estimates is similar to those in Columns (1)-(3), but the magnitudes of estimates on the distance indicators and gender-interacted distances are higher because the sample mean of the dependent variable for adoption within 10 years are more than double that for adoption within 5 years. At the same time, the estimates on the three-way interactions are largely unchanged, suggesting that the gender-match effect is attenuated (in percentage terms) in the longer time window. This is intuitive given that as time passes, the underlying value of a new idea is likely better appreciated by researchers so that adoption is less affected by the demographic characteristics of the innovators or adopters.

3.3 Potential mechanisms

Researchers have long argued that there is a “Matthew effect” in science (Merton, 1968), whereby the most prominent scientists receive the most attention for their work. If so, new ideas introduced by more prominent researchers may receive more attention than those introduced by less prominent researchers and, if women are, on average, less prominent than men, this disparity may help to explain lower adoption of ideas introduced by women. Put differently, our findings

¹⁵ The time window of available years of the Scimago journal rank (1999-2021) does not align perfectly with the time window for the birth of new ideas (1980-2008). We take an average over the years for which we have the journal ranking to measure journal rank.

may reflect people tending to adopt ideas introduced by more prominent innovators, rather than a gender difference *per se*. To gauge this possibility, we construct a measure of a researcher's prominence using the number of publications weighted by journal rank. Formally, we generate a high-prominence innovator dummy (*HI*) equal to one if the closest innovator to a potential adopter is above the median in the distribution of innovator productivity on this measure. The results are reported in Table 4. When the prominence measure is included along with interactions with distance, there is some evidence for a Matthew effect (e.g., at distance 2), but the estimates are noisy. When we include interactions between this prominence measure, innovator gender, and distance, we find that ideas introduced by prominent women have higher adoption than those introduced by less prominent women at distance 1. At the same time, even conditional on the prominence of innovators, there still exists a remaining gender gap in adoption rates when men are closest to male and female innovators.

Next, we explore whether the gender effects differ across female- versus male-dominated research fields. In the context of venture capital funding for entrepreneurs, Herbert (2020) shows that investors are more likely to invest in female-founded start-ups in female-dominated sectors. The primary reason is that the abilities of entrepreneurs in gender-incongruent sectors are underestimated. Similar forces may operate in our setting, with the ideas of female innovators being undervalued in male-dominated fields.¹⁶

To investigate possible differences across fields, we constructed the proportion of female researchers in each field. To do this, we rely on Medical Subject Headings (MeSH), a curated classification system for biomedical articles. MeSH codes are assigned to articles in our data by

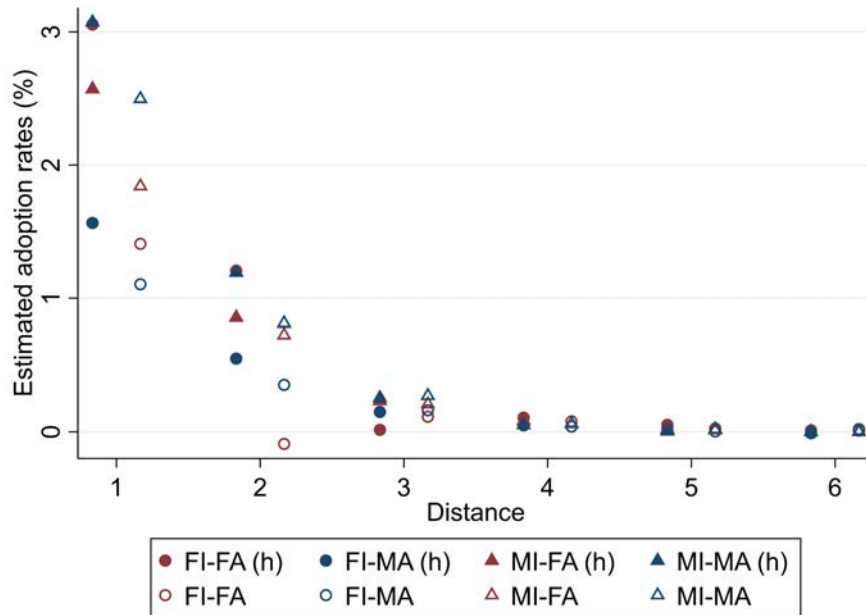
¹⁶ We are thankful to the anonymous reviewer for providing this insightful and helpful suggestion.

independent experts at the National Library of Medicine.¹⁷ Specifically, we take the share of women among the authors on all articles in each MeSH category. As articles are frequently tagged with multiple MeSH terms, when calculating the gender mix of fields, we include articles in all of the MeSH categories associated with an article. When calculating the relevant gender mix for a new idea, we average the share of women across all of the MeSH categories associated with the articles that introduce that idea. This share of women among authors in the “field” where a new idea was born ranges from 0.22 to 0.46.

We divide the sample at the median, which is 0.35, and estimate the main specification using both subsamples. Figure 6 shows the estimates. The solid shapes represent ideas born in fields with more female researchers and the hollow shapes represent ideas born in fields with fewer female researchers. There are two things worth noting. First, men lag in the adoption of women's ideas at short distances in both cases. Second, women are more likely to adopt new ideas introduced by female innovators in fields with more women. In Herbert's framework, the heterogenous effects for women but not for men is consistent with men exercising taste-based discrimination against ideas introduced by women, while women's use of new ideas is more context-based.

¹⁷ The MeSH terms have different levels representing broader or narrower categories. We use level-3 MeSH terms which have a total of 1517 terms. For each MeSH term, we calculate the share of female researchers who have published research articles tagged with that term.

Figure 6. Stratifying by the Gender Composition of Fields

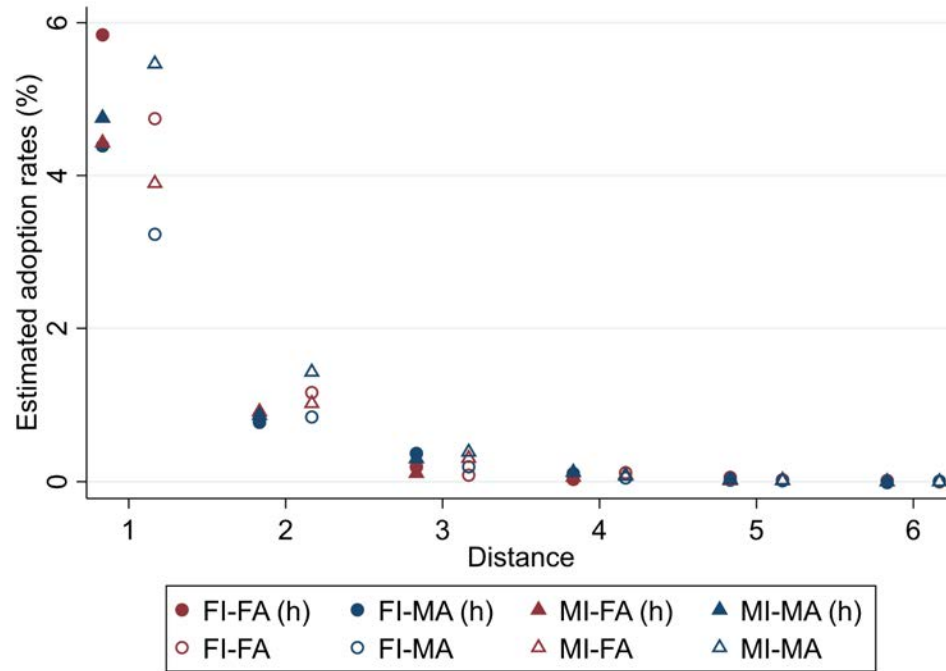


Notes: The figure plots regression-based adoption rates for different combinations of the gender of a potential adopter and his or her closest innovator stratified by the gender mix of the field in which the idea was introduced. Fields with an above median share of women are indicated by (h). *FA / MA* indicate a female / male potential adopter. *FI / MI* indicate that the closest innovator is female / male respectively. The estimates used to generate this figure are reported in Table A1 in the appendix.

Finally, the prestige or visibility of the journal in which an idea is published may affect its adoption and may reflect underlying differences in the quality of the idea. To address these possibilities, we estimate heterogeneous effects according to the ranking of the journal(s) that first published the new ideas. We contrast the important new ideas published in the top quartile of journals to the other important new ideas in our sample and estimate the regression separately on the two subsamples. Figure 7 shows the estimates, where the solid markers represent new ideas first published in higher-ranked journals and the hollow markers represent those published in lower-ranked journals. There are three findings: first, at short distances, male potential adopters are less likely to adopt new ideas introduced by female innovators (FI-MA blue circles) regardless of journal rank. Second, female potential adopters are disproportionately more likely to adopt new ideas that were introduced by nearby female innovators and were first published in better journals.

Last but not least, the solid symbols are less dispersed than the hollow symbols. This implies that new ideas published by top journals are less affected by gender.

Figure 7. Estimates by Journal Rank



Notes: The figure plots regression-based adoption rates for different combinations of the gender of a potential adopter and his or her closest innovator stratified by the ranking of the journal which first published the new idea. New ideas published by journals ranked in the top quartile are indicated by (h). FA / MA indicate a female / male potential adopter. FI / MI indicate that the closest innovator is female / male respectively. The estimates used to generate this figure are reported in Table A2 in the appendix.

3.4 Decomposition analysis

To gauge the relative importance of factors in the lower adoption rates of women’s new ideas, we conduct a series of decomposition exercises. Our first decomposition focuses on the adoption of ideas based on the gender of innovators. The second set of two decompositions studies the gender of potential innovators and potential adopters. A fourth decomposition focuses on the gender mix of the entire team of innovators.

For the first counterfactual scenario, we explore the effects of differences in the distribution of network distances between female innovators and male innovators ignoring the genders of potential adopters. Specifically, we use the model,

$$A_{i,j} = \alpha_0 + \sum_d \alpha_d \cdot 1(\text{Distance}_{i,j} = d) + \sum_d \eta_d \cdot 1(\text{Distance}_{i,j} = d) \cdot X_{i,j}^I + \gamma_i + \rho_j + u_{i,j} \quad (2)$$

Let $X_{i,j}^I = 1$ indicate that the innovator on idea i who is closest to potential adopter j is female. After partialling out the idea and adopter fixed effects, the adoption rates at each distance implied by the regression are $R_d^{MI} = \alpha_d$ and $R_d^{FI} = \alpha_d + \eta_d$. Let P_d^{MI} and P_d^{FI} denote the proportions of researchers at distance d from male and female innovators, respectively. The implied adoption rates are $\sum_d P_d^{MI} \times R_d^{MI}$ and $\sum_d P_d^{FI} \times R_d^{FI}$, respectively. Keeping the adoption rates for MI (R_d^{MI}) and FI (R_d^{FI}) at each distance unchanged, we calculate the change in the adoption rate experienced by female innovators if they had the same distribution of network distances as male innovators. Formally, we calculate $P_d^{MI} \times R_d^{FI} - P_d^{FI} \times R_d^{FI}$. We then divide this gap by the overall gap in adoption rates between FI and MI, $\sum_d (P_d^{MI} \times R_d^{MI} - P_d^{FI} \times R_d^{FI})$, to obtain the share of the total gap that can be explained by each distance. Given that the estimated coefficients are quite imprecise at greater distances, and that the adoption rates are highest at distances 1 and 2, we focus on distances 1, 2, and combine distances greater than 2 for the decomposition analysis.¹⁸

Table 5 shows the share of the total difference in the adoption rate of the ideas of female innovators relative to male innovators accounted for by each difference by distance and across all

¹⁸ A decomposition using estimates from including all six distance categories yields similar results.

distances (overall). Column 1 shows that if FIs had the same network connections as MIs¹⁹, but that the adoption rates remained unchanged at each distance, the gender gap in adoption would be reduced by 24%. Women have fewer potential adopters at close distances and hence more at long distances. Thus, giving FIs the network connections of MIs increases the adoption of FIs' ideas. Nevertheless, 76% of the gap in adoption between FIs and MIs is due to a lower adoption rate for the ideas of FIs compared to MIs at each distance.

Next, we turn to models of adoption that also consider the gender of potential adopters. There are two natural counterfactuals to consider – we can (A) give women the same gender mix of potential adopters as men or (B) give women the same mix of same-gender adopters as men. The estimates used in this exercise are from the estimation of Eq (1). Let $X_j^{PA}=1$ indicate that potential adopter j is female. Table 6 summarizes the predicted adoption rates for ideas introduced by women and men (after partialling out the idea and potential adopter fixed effects) based on Eq (1).

Let G_d^{MI} and G_d^{FI} represent the proportion of women among all potential adopters at distance d from male and female innovators, respectively. Then, at distance d , the predicted adoption rate for FI is $P_d^{FI} \times R_d^{FIFA} \times G_d^{FI} + P_d^{FI} \times R_d^{FIMA} \times (1 - G_d^{FI})$, and that for MI is $P_d^{MI} \times R_d^{MIFA} \times G_d^{MI} + P_d^{MI} \times R_d^{MIMA} \times (1 - G_d^{MI})$. We first give FIs the gender mix of potential adopters for MIs at each distance (i.e., replace FI's proportion of female researchers at each distance, G_d^{FI} , with its male counterpart, G_d^{MI}) but keep unchanged the distribution of distances and the adoption rates for all four combinations FI-FA, MI-FA, FI-MA, MI-MA at each distance. The results in Column (2A) of Table 5 indicate that the adoption rate of FIs' ideas is actually somewhat lower. Intuitively, for FIs, replacing FAs with MAs reduces adoption at short distances

¹⁹ We obtain the distribution of network distance for male innovators by restricting to new ideas with no female innovators.

because, despite the overall higher adoption rate of MAs, at short distances, FAs have higher adoption rates for FIs than MAs. Yet, the overall effects are relatively small because the difference in the gender composition of potential adopters is not that great.

Because we find that the gender match between innovators and potential adopters plays an important role, we explore an alternative counterfactual using this model. Namely, we consider giving FIs as many *same-gender* potential adopters as MIs (i.e. giving FIs as many FAs as MIs have MAs while preserving the total number of potential adopters and adoption rates of each gender pair at each distance). Because most potential adopters are men, this counterfactual has a much larger effect on the gender mix of potential adopters than the previous counterfactual. Intuitively, at short distances, FAs have much higher adoption rates than MAs for FIs, so increasing the number of FAs at short distances increases adoption for FIs. Overall, giving FIs the same share of same-gender adopters at each distance as MIs would reduce the gender gap by 22%.

Our main findings from these decompositions are that the network positions of women and lower overall adoption for female innovators account respectively for 24% and 76% of the lower adoption rates for female innovators. Moreover, the facts that there are more men than women in biomedicine and that people tend to adopt ideas introduced by people of their own gender further reduces the adoption of ideas introduced by women by roughly 20%.

Our fourth decomposition returns to the analysis of teams from the introduction. Specifically, we study the gap in adoption between female-majority and male-majority ideas. New ideas introduced by teams with over 50% men have an adoption rate that is 25% higher than ideas by female majority teams (0.1% vs 0.08%). By construction, people are more likely to be close to an FI than an MI on female majority ideas, but the two are not identical (because many inventor teams include men and women). How much does the lower adoption rate of FIs compared with MIs

translate into differences in adoption rates of male majority ideas and female majority ideas? To answer this question, we first limit the sample to one in which each potential adopter is closest to exactly one innovator so that the gender of the closest innovator is that of a single person rather than an average of multiple innovators²⁰. Then we consider two hypothetical new ideas that represent respectively an average female majority idea and an average male majority idea. Both ideas have 5 innovators, which is the average number of innovators associated with a new idea. The hypothetical male majority idea has 12.1% (0.6) women innovators on the team; the hypothetical female majority idea has 65.6% (3.3) women innovators on the team. By using the sample average distribution of network connections for MIs and FIs and the adoption rates at each distance given the gender of potential adopters at each distance in Table 6, we find lower adoption rates for FIs explain 51% of the difference in adoption rates for the hypothetical male majority idea and the hypothetical female majority idea.²¹

4 Robustness checks

1) Different roles of researchers in an innovator team

In the above, we treat each author on articles equally. However, in practice, the positions of authors on biomedical research articles provide information about their contribution to the works (i.e., author order is rarely assigned alphabetically). Specifically, the first author is typically the person who had primary responsibility for the work and the last author is generally the senior author (e.g., the person who ran the lab in which the work was conducted) and the first and last

²⁰ This subsample takes up 60% of the full sample.

²¹ It is worth noting that this number may not apply exactly to the full sample, but it still suggests that less adoption for FI ideas is an important reason behind the lower adoption rate of female majority ideas.

authors typically receive the most credit. One might be concerned that the gender gap in adoption rates we estimate is affected by our inclusion of middle authors who may have made relatively modest contributions to the work and are not closely associated with it in the minds of potential adopters.

We address the possibility that gender affects adoption differently by author position in two ways. First, we assume that a new idea is attributed only to the first and last authors (on any article that introduces the idea) and repeat the analysis. The estimates are reported in Table 7, and they are similar to those in Table 3, although greater adoption of the ideas of female innovators by female adopters is restricted to distance 1.

Second, we test whether the gender of the first/last author has a larger effect on adoption than the gender of middle authors. We include two sets of three-way interactions in our regression: “*distance* × *nearest to female first/last-author innovator* × *female potential adopter*” and “*distance* × *nearest to female middle-author innovator* × *female potential adopter*”. The estimates are reported in Table 8. To facilitate comparison, we put the estimates side by side. First, both female middle authors and female first and last authors have lower adoption rates. Second, the pattern of gender match effects for innovators is qualitatively similar for first/last-authors and middle-authors. However, the magnitude of estimates is larger for the gender of first/last-author innovators.

2) Names with Uncertain / Ambiguous Genders

Our gender variable is based on predictions using names, but some names are strongly predictive of gender while others are ambiguous. This is true for many English names and even more true for

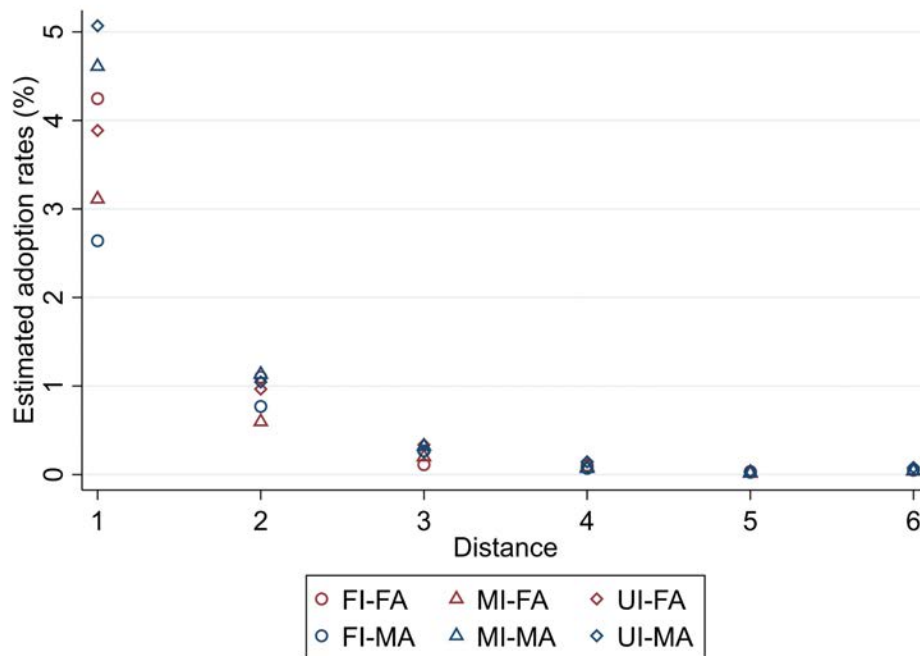
Asian names when written in Latin / Roman characters rather than in their original characters. We address this issue in two ways.

First, we exclude innovators with Asian names. Over 48% of all Asian innovators have uncertain / ambiguous gender predictions, compared to 12% for other racial / ethnic groups. The estimates dropping people with Asian names are very similar to those from the full sample.

Second, we take advantage of the continuous, predicted probabilities that a name belongs to a woman or man provided by Genni (Smith et al., 2013). Specifically, we divide our sample into three categories: 1) probability of the closest innovator being a female is above 95% (FI); 2) probability of closest innovator being a male is above 95% (MI); and 3) cases where the closest innovator has an uncertain / ambiguous gender prediction defined as between 5% and 95% female (UI)²². We exclude cases where a researcher is equidistant from multiple innovators so that the gender of the closest innovator is based on a single person. As in our main specification, we estimate the three-way interactions between distance, gender of the closest innovator, and gender of the potential adopter. Here, however, instead of a binary variable (woman or not), we now have three categories, so we use two dummy variables to represent innovator gender, and then interact both with distance and the (binary) gender of the potential adopter. Figure 8 shows the estimates. The adoption of ideas from innovators of uncertain / ambiguous gender (UI) is represented by diamonds. Compared with ideas introduced by women (circles), the adoption rates of UI are higher. Men do not seem to have lower adoption rates for ideas of UI in the way that they do for new ideas from innovators (FI). These findings suggest that the gender effects are concentrated on innovators with gender-specific names.

²² 17.4% of all observations in the sample are in the UI category.

Figure 8. Estimates that account for gender uncertain / ambiguous names.



Notes: The figure plots regression-based adoption rates for different combinations of the gender of a potential adopter and his or her closest innovator. *FA* and *MA* indicate a female or male potential adopter. *FI*, *MI*, and *UI* indicate that the closest innovator is female, male, or has an uncertain / ambiguous gender respectively.

3) Multiple innovator teams

One issue with our estimates is that when multiple teams introduce the same idea in the same year, it is possible that both teams developed the idea independently or that one team developed the idea first and the other adopted it quickly. To address this issue, Appendix Table A3 provides estimates for ideas introduced by single teams. The estimates are broadly similar to our main estimates.

4) Redefining concepts

As the availability of abstracts in Medline has greatly improved over the years, the birth year of some ideas may be miscalculated if they appear first in abstracts but were not covered in Medline.

Therefore, we repeat our analysis using only paper titles and report the estimates in Table 9 Column (1). The results are qualitatively similar to our analysis using both titles and abstracts, although the magnitudes of the estimates are smaller due to lower adoption rates when we limit only to the text in paper titles.

Our main estimates identify the first appearance of a word or a word combination as a new idea. One shortcoming of this definition is that synonyms are treated as different ideas. Although we have no reason to believe that men or women are more likely to introduce or use synonyms, which would affect our estimates, we conduct a robustness analysis using a different definition of ideas/concepts based on the Unified Medical Language Systems® (UMLS®). The UMLS brings together many health and biomedical vocabularies and standards to make it possible to extract concepts from text. It provides all the language terms that are associated with a concept.²³ Following Packalen and Bhattacharya (2020), we extract all the UMLS concepts from titles of publications. When ranking all new concepts within a cohort by the number of mentions, we take the top 0.5% of all UMLS concepts. The sample construction is the same as explained in Section 2. The estimates of Eq (1) using this new sample are reported in Table 9 Column (2).

4. Conclusion

This paper studies the diffusion of new ideas through networks and how diffusion is related to the gender, race and ethnicity of innovators and potential adopters. Network proximity to innovators of a new idea is associated with a higher rate of adoption. Using a novel, two-way fixed

²³ For example, the concept of HIV has a concept id of C0019682, and it corresponds to many “atoms”. Each atom is a unique concept name or string from a source vocabulary. For example, the atoms associated with HIV include AIDS Virus, Human Immunodeficiency Viruses, Virus-HIV, and many others.

effects strategy, we explore two reasons why ideas introduced by female scientists are under-utilized. First, female innovators are not as well-connected as men. Second, even at short distances, people are less likely to adopt ideas introduced by women, and this effect is particularly true for male potential adopters.

Although our analysis has focused on gender, we find concerning patterns for non-Hispanic Blacks and Hispanics, where their ideas are less likely to be adopted compared with non-Hispanic Whites. Our analysis suggests that this gap is not entirely because that they are more disadvantaged in terms of network positions. A similar analysis might also be conducted to understand disparities across age groups.

Table 1 Summary Statistics

	<i>N</i>	<i>Mean</i>	<i>SD</i>
<i>Panel A. Innovators</i>			
<i>Female</i>	21,822	0.25	0.43
<i>Male</i>	21,822	0.56	0.50
<i>Asian</i>	21,822	0.16	0.37
<i>Black</i>	21,822	0.006	0.077
<i>Hispanic</i>	21,822	0.037	0.188
<i>White</i>	21,822	0.80	0.40
<i>Career age</i>	21,822	8.15	8.77
<i>Panel B. Potential adopters</i>			
<i>Female</i>	536,813	0.33	0.47
<i>Male</i>	536,813	0.52	0.50
<i>Asian</i>	536,813	0.13	0.33
<i>Black</i>	536,813	0.006	0.08
<i>Hispanic</i>	536,813	0.04	0.20
<i>White</i>	536,813	0.83	0.38
<i>Panel C. Potential adopter-idea pairs</i>			
<i>Adoption within 5 years</i>	7,600,166	0.10%	3.47%
<i>Adoption within 10 years</i>	7,600,166	0.25%	3.14%
<i>Network distance</i>	7,600,166	6.16	1.61
<i>Career age</i>	7,600,166	16.28	11.05

Notes: Each observation is an idea-innovator in Panel A, a unique potential adopter in Panel B and a potential adopter-idea pair in Panel C. In Panel A, it is possible that one innovator may be associated with multiple new ideas. Adoption within 5 years (10 years) is a binary measure of whether a researcher adopts a new idea within 5 /10 years after the idea is born. The mean and standard deviation of adoption are multiplied by 100 so that they are in percentage points.

Table 2 Baseline regressions

	Adoption		
	(1)	(2)	(3)
<i>Distance=1</i>	2.47215***	2.59181***	2.48191***
	(0.1726)	(0.0372)	(0.0372)
<i>Distance=2</i>	0.99258***	0.96669***	0.88323***
	(0.0509)	(0.0167)	(0.0168)
<i>Distance=3</i>	0.30915***	0.29104***	0.24454***
	(0.0133)	(0.0076)	(0.0079)
<i>Distance=4</i>	0.09282***	0.07907***	0.06114***
	(0.0046)	(0.0043)	(0.0047)
<i>Distance=5</i>	0.02662***	0.01882***	0.01480***
	(0.0026)	(0.0032)	(0.0034)
<i>Characteristics of the Closest Innovators</i>			
<i>Female</i>	-0.01740***	-0.02015***	-0.00885
	(0.0035)	(0.0038)	(0.0060)
<i>Asian</i>	-0.00763	-0.01222**	0.00847
	(0.0047)	(0.0048)	(0.0077)
<i>Black</i>	-0.04866***	-0.05160**	-0.03673
	(0.0166)	(0.0211)	(0.0300)
<i>Hispanic</i>	-0.01936**	-0.00964	0.00707
	(0.0088)	(0.0092)	(0.0131)
<i>Career age</i>	-0.00594***	-0.00508***	0.00098
	(0.0004)	(0.0004)	(0.0007)
<i>Career age squared</i>	0.00007***	0.00006***	-0.00004**
	(0.0000)	(0.0000)	(0.0000)
<i>Characteristics of potential adopters</i>			
<i>Female</i>	-0.02916***		
	(0.0026)		
<i>Asian</i>	-0.01277***		
	(0.0039)		

<i>Black</i>	-0.01546		
	(0.0140)		
<i>Hispanic</i>	-0.01264*		
	(0.0065)		
<i>Career age</i>	-0.00071**		
	(0.0003)		
<i>Career age squared</i>	-0.00004***		
	(0.0000)		
<i>Individual FEs</i>	No	Yes	Yes
<i>Idea FEs</i>	No	No	Yes
<i>N</i>	7,600,166	7,497,421	7,497,382
<i>R squared</i>	0.002	0.041	0.047

Notes: Each observation is an idea-potential adopter pair. The outcome variable is a binary variable of adoption or not within 5 years after a new idea was introduced. All estimated coefficients are multiplied by 100.

Table 3 Gender and network distance

	Adoption within 5 years			Adoption within 10 years		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Distance=1</i>	2.61543***	2.73366***	2.61122***	4.19580***	4.48938***	4.48210***
	(0.0401)	(0.0461)	(0.0610)	(0.0633)	(0.0727)	(0.0961)
<i>Distance=2</i>	0.94209***	0.98867***	0.98898***	1.59183***	1.68450***	1.70389***
	(0.0181)	(0.0207)	(0.0276)	(0.0285)	(0.0326)	(0.0434)
<i>Distance=3</i>	0.26076***	0.27333***	0.29177***	0.50675***	0.53171***	0.56786***
	(0.0084)	(0.0096)	(0.0128)	(0.0132)	(0.0151)	(0.0201)
<i>Distance=4</i>	0.06121***	0.05858***	0.06078***	0.13867***	0.13835***	0.13424***
	(0.0049)	(0.0056)	(0.0074)	(0.0078)	(0.0088)	(0.0116)
<i>Distance=5</i>	0.01453***	0.01341***	0.00921*	0.02650***	0.02356***	0.00704
	(0.0036)	(0.0041)	(0.0054)	(0.0057)	(0.0065)	(0.0084)
<i>FI*Distance=1</i>	0.00391	0.00177	-1.38880***	0.00350	-0.00295	-2.09490***

	(0.0072)	(0.0078)	(0.1320)	(0.0114)	(0.0123)	(0.2078)
<i>FI*Distance=2</i>	-0.97823***	-1.39575***	-0.55993***	-1.60678***	-1.96524***	-0.78777***
	(0.1071)	(0.1295)	(0.0570)	(0.1689)	(0.2042)	(0.0898)
<i>FI*Distance=3</i>	-0.42494***	-0.51529***	-0.09980***	-0.56024***	-0.74611***	-0.16517***
	(0.0475)	(0.0560)	(0.0262)	(0.0749)	(0.0882)	(0.0413)
<i>FI*Distance=4</i>	-0.12784***	-0.11050***	-0.01153	-0.15140***	-0.16538***	-0.05693**
	(0.0219)	(0.0257)	(0.0150)	(0.0346)	(0.0406)	(0.0237)
<i>FI*Distance=5</i>	-0.00482	-0.01217	-0.00162	-0.04994**	-0.05884**	-0.02465
	(0.0127)	(0.0147)	(0.0116)	(0.0200)	(0.0232)	(0.0183)
<i>FI</i>	0.00165	-0.00272	0.00119	-0.00870	-0.02216	-0.00492
	(0.0098)	(0.0113)	(0.0080)	(0.0155)	(0.0179)	(0.0126)
<i>FA*Distance=1</i>		-0.48612***	-0.58752***		-1.20685***	-1.31566***
		(0.0934)	(0.0949)		(0.1473)	(0.1494)
<i>FA*Distance=2</i>		-0.19708***	-0.20645***		-0.39206***	-0.38326***
		(0.0422)	(0.0428)		(0.0666)	(0.0673)
<i>FA*Distance=3</i>		-0.05153***	-0.05750***		-0.10248***	-0.11392***
		(0.0188)	(0.0190)		(0.0297)	(0.0300)
<i>FA*Distance=4</i>		0.01050	0.01233		0.00127	0.00328
		(0.0105)	(0.0106)		(0.0165)	(0.0167)
<i>FA*Distance=5</i>		0.00446	0.00432		0.01155	0.01090
		(0.0076)	(0.0078)		(0.0121)	(0.0122)
<i>FA*FI</i>		0.00748	0.00898		0.02243	0.02524
		(0.0106)	(0.0108)		(0.0167)	(0.0171)
<i>FA*FI*Distance=1</i>		1.40141***	1.23405***		1.40137***	1.37333***
		(0.2314)	(0.2360)		(0.3649)	(0.3715)
<i>FA*FI*Distance=2</i>		0.34860***	0.41633***		0.71406***	0.79618***
		(0.1054)	(0.1075)		(0.1663)	(0.1692)
<i>FA*FI*Distance=3</i>		-0.05312	-0.05676		0.06365	0.04755
		(0.0479)	(0.0490)		(0.0756)	(0.0771)
<i>FA*FI*Distance=4</i>		0.02498	0.02139		0.03183	0.02487
		(0.0267)	(0.0273)		(0.0421)	(0.0431)

<i>FA*FI*Distance=5</i>		0.01587	0.01366		0.04900	0.05234
		(0.0207)	(0.0212)		(0.0327)	(0.0334)
<i>Journal Rank**Distance=1</i>			0.02872***			0.01123
			(0.0085)			(0.0133)
<i>Journal Rank**Distance=2</i>			0.00203			-0.00485
			(0.0042)			(0.0067)
<i>Journal Rank**Distance=3</i>			-0.00414**			-0.00833***
			(0.0020)			(0.0032)
<i>Journal Rank**Distance=4</i>			-0.00057			0.00090
			(0.0011)			(0.0017)
<i>Journal Rank**Distance=5</i>			0.00105			0.00379***
			(0.0007)			(0.0012)
<i>Individual FEs</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Idea FEs</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Career age of closest innovator</i>	No	No	Yes	No	No	Yes
<i>N</i>	7,497,382	7,497,382	7,303,768	7,497,382	7,497,382	7,303,768
<i>R squared</i>	0.047	0.047	0.047	0.049	0.049	0.050

Notes: The outcome variable for Columns (1)-(3) is a binary variable for adoption within 5 years after the birth of a new idea, and that for Columns (4)-(6) is a binary variable for adoption within 10 years after the birth of a new idea. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. Journal Rank is the Scimago journal rank of the journal that first published the new idea. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects. Columns (3) and (6) additionally control for the career age of the closest innovator.

Table 4 Prominence of innovators

	Adoption
<i>Distance=1</i>	2.80239***
	(0.0659)

<i>Distance=2</i>	0.85662***
	(0.0304)
<i>Distance=3</i>	0.29926***
	(0.0138)
<i>Distance=4</i>	0.06134***
	(0.0077)
<i>Distance=5</i>	0.01631***
	(0.0054)
<i>FI*Distance=1</i>	-0.48251***
	(0.0935)
<i>FI*Distance=2</i>	-0.20295***
	(0.0422)
<i>FI*Distance=3</i>	-0.05059***
	(0.0188)
<i>FI*Distance=4</i>	0.01050
	(0.0105)
<i>FI*Distance=5</i>	0.00449
	(0.0076)
<i>FI</i>	-0.00449
	(0.0100)
<i>FA*Distance=1</i>	-1.88554***
	(0.1772)
<i>FA*Distance=2</i>	-0.42496***
	(0.0775)
<i>FA*Distance=3</i>	-0.13286***
	(0.0347)
<i>FA*Distance=4</i>	-0.01367
	(0.0192)
<i>FA*Distance=5</i>	-0.01359
	(0.0143)
<i>FA*FI</i>	0.00745
	(0.0106)
<i>FA*FI*Distance=1</i>	1.45533***
	(0.2319)
<i>FA*FI*Distance=2</i>	0.35805***
	(0.1055)
<i>FA*FI*Distance=3</i>	-0.05425
	(0.0479)
<i>FA*FI*Distance=4</i>	0.02485

	(0.0267)
<i>FA*FI*Distance=5</i>	0.01602
	(0.0207)
<i>High-productivity innovator (HI)</i>	-0.00787
	(0.0055)
<i>HI*Distance=1</i>	-0.11878
	(0.0814)
<i>HI*Distance=2</i>	0.21642***
	(0.0367)
<i>HI*Distance=3</i>	-0.04235***
	(0.0164)
<i>HI*Distance=4</i>	-0.00467
	(0.0091)
<i>HI*Distance=5</i>	-0.00522
	(0.0067)
<i>HI*FI</i>	0.01204
	(0.0141)
<i>HI*FI*Distance=1</i>	0.87136***
	(0.2156)
<i>HI*FI*Distance=2</i>	-0.14166
	(0.0957)
<i>HI*FI*Distance=3</i>	0.03457
	(0.0441)
<i>HI*FI*Distance=4</i>	0.00144
	(0.0254)
<i>HI*FI*Distance=5</i>	0.02316
	(0.0198)
<i>N</i>	7,497,382
<i>R squared</i>	0.047

Notes: The outcome variable is a binary variable for adoption within 5 years after a new idea is born. *High-productivity innovator (HI)* is a binary indicator of whether the closest innovator has above median (Column 1) or top 25% (Column 2) productivity among all innovators. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects.

Table 5. Decompositions of Gender Differences in Adoption

Difference in Adoption of FI relative to MI by Distance:	1) Differences in the distribution of network distances between female and male innovators	2) Counterfactual scenarios for the gender of potential adopters, by distance	
		2A) FIs have the same gender mix as MI	2B) FIs have the same gender-match as MI
1	6.0%	-2.37%	9.06%
2	9.5%	-1.26%	9.27%
>=3	8.2%	-0.39%	3.56%
Overall	23.70%	-4.02%	21.90%

Notes: This table shows three decomposition analyses. The numbers show how much of the total gap in adoption rates of women's and men's ideas are due to the indicated differences at each distance.

Table 6 Predicted adoption rates

Innovator-adopter gender combination	Predicted adoption rates at distance d (R_d)
MI-MA	α_d
MI-FA	$\alpha_d + \beta_d$
FI-MA	$\alpha_d + \eta_d$
FI-FA	$\alpha_d + \beta_d + \eta_d + \lambda_d$

Notes: This table provides the predicted adoption rates for different gender combinations by Eq (1), after partialling out the idea and potential adopter fixed effects.

Table 7 Restricting to first and last author innovators

	Adoption		
	(1)	(2)	(3)
<i>Distance=1</i>	3.27001***	3.58556***	3.77103***
	(0.0525)	(0.0606)	(0.0807)
<i>Distance=2</i>	1.22453***	1.33658***	1.23477***
	(0.0246)	(0.0281)	(0.0380)
<i>Distance=3</i>	0.35124***	0.36296***	0.33434***
	(0.0113)	(0.0129)	(0.0174)
<i>Distance=4</i>	0.10169***	0.09297***	0.07713***
	(0.0065)	(0.0073)	(0.0097)
<i>Distance=5</i>	0.02351***	0.01637***	0.01058
	(0.0046)	(0.0052)	(0.0069)
<i>FI*Distance=1</i>	-0.00395	-0.00290	-0.00464
	(0.0096)	(0.0102)	(0.0104)
<i>FI*Distance=2</i>	-1.79177***	-2.57866***	-2.61440***
	(0.1436)	(0.1704)	(0.1736)
<i>FI*Distance=3</i>	-0.00168	0.10307	0.09241
	(0.0668)	(0.0786)	(0.0803)
<i>FI*Distance=4</i>	-0.11697***	-0.13685***	-0.12798***
	(0.0305)	(0.0355)	(0.0362)
<i>FI*Distance=5</i>	-0.01205	-0.01793	-0.01791
	(0.0171)	(0.0197)	(0.0202)
<i>FI</i>	-0.00209	-0.00969	-0.00977
	(0.0126)	(0.0145)	(0.0149)
<i>FA*Distance=1</i>		-1.27082***	-1.28382***
		(0.1215)	(0.1235)
<i>FA*Distance=2</i>		-0.47603***	-0.44637***
		(0.0575)	(0.0585)
<i>FA*Distance=3</i>		-0.04864*	-0.04194
		(0.0257)	(0.0261)
<i>FA*Distance=4</i>		0.03604***	0.03460**
		(0.0139)	(0.0141)
<i>FA*Distance=5</i>		0.02910***	0.02911***
		(0.0098)	(0.0100)
<i>FA*FI</i>		-0.00365	-0.00570
		(0.0127)	(0.0129)
<i>FA*FI*Distance=1</i>		2.86759***	2.98785***

		(0.3169)	(0.3231)
<i>FA*FI*Distance=2</i>		-0.28012*	-0.42396***
		(0.1484)	(0.1522)
<i>FA*FI*Distance=3</i>		0.08042	0.05847
		(0.0676)	(0.0690)
<i>FA*FI*Distance=4</i>		0.01841	0.01909
		(0.0362)	(0.0370)
<i>FA*FI*Distance=5</i>		0.02510	0.02652
		(0.0265)	(0.0270)
<i>Individual FE</i>	Yes	Yes	Yes
<i>Idea FE</i>	Yes	Yes	Yes
<i>Journal rank * Distance FEs</i>	No	No	Yes
<i>N</i>	5,767,175	5,767,175	5,605,960
<i>R squared</i>	0.055	0.055	0.056

Notes: The outcome variable is a binary variable for adoption within 5 years after a new idea is born. The innovators of a new idea are the first and last authors of papers that first introduced the new idea, but not the middle authors. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. Journal rank is the Scimago journal rank of the journal that first published the new idea. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects.

Table 8 Middle authors versus first and last authors

Adoption			
<i>Distance=1</i>	2.77478***	<i>FA*Distance=1</i>	-0.73907***
	(0.0628)		(0.1027)
<i>Distance=2</i>	1.06790***	<i>FA*Distance=2</i>	-0.21139***
	(0.0285)		(0.0462)
<i>Distance=3</i>	0.29886***	<i>FA*Distance=3</i>	-0.03465*
	(0.0133)		(0.0207)
<i>Distance=4</i>	0.06802***	<i>FA*Distance=4</i>	0.00817
	(0.0077)		(0.0117)
<i>Distance=5</i>	0.01687***	<i>FA*Distance=5</i>	0.00566
	(0.0057)		(0.0087)
<i>Mid_FI*Distance=1</i>	-1.16134***	<i>Last_FI*Distance=1</i>	-1.41962***
	(0.1435)		(0.1156)
<i>Mid_FI*Distance=2</i>	-0.48060***	<i>Last_FI*Distance=2</i>	-0.62553***

	(0.0606)		(0.0505)
<i>Mid_FI*Distance=3</i>	-0.00296	<i>Last_FI*Distance=3</i>	-0.13173***
	(0.0263)		(0.0224)
<i>Mid_FI*Distance=4</i>	-0.01629	<i>Last_FI*Distance=4</i>	-0.04082***
	(0.0145)		(0.0124)
<i>Mid_FI*Distance=5</i>	-0.00959	<i>Last_FI*Distance=5</i>	-0.03785***
	(0.0109)		(0.0090)
<i>FA*Mid_FI</i>	-0.00149	<i>FA*Last_FI</i>	0.00893
	(0.0103)		(0.0082)
<i>FA*Mid_FI*Distance=1</i>	0.48318*	<i>FA*Last_FI*Distance=1</i>	1.47611***
	(0.2643)		(0.2109)
<i>FA*Mid_FI*Distance=2</i>	0.19113*	<i>FA*Last_FI*Distance=2</i>	0.28306***
	(0.1161)		(0.0964)
<i>FA*Mid_FI*Distance=3</i>	-0.06693	<i>FA*Last_FI*Distance=3</i>	-0.09594**
	(0.0492)		(0.0421)
<i>FA*Mid_FI*Distance=4</i>	0.03511	<i>FA*Last_FI*Distance=4</i>	0.01275
	(0.0263)		(0.0224)
<i>FA*Mid_FI*Distance=5</i>	0.00421	<i>FA*Last_FI*Distance=5</i>	0.00190
	(0.0197)		(0.0164)
<i>N</i>	7,303,768	<i>R-sq</i>	0.048

Notes: This table shows the estimates from one regression. *Last_FI* means the closest innovator is a woman and the first or last author on the paper which first introduced a new idea. *Mid_FI* means the closest innovator is a woman and a middle author. Individual fixed effects and idea fixed effects are controlled. All estimated coefficients are multiplied by 100.

Table 9 Alternative definitions of innovations

	Ideas in titles only	UMLS concepts
	(1)	(2)
<i>Distance=1</i>	0.53981***	6.09808***
	(0.1207)	(0.2355)
<i>Distance=2</i>	0.75219***	2.03450***
	(0.0482)	(0.0969)
<i>Distance=3</i>	0.23693***	0.37767***
	(0.0216)	(0.0462)
<i>Distance=4</i>	0.06348***	0.08132***
	(0.0124)	(0.0281)
<i>Distance=5</i>	0.01519*	-0.00302

	(0.0091)	(0.0215)
<i>FI</i>	-0.00203	-0.00745
	(0.0136)	(0.0230)
<i>FI*Distance=1</i>	-0.93673***	-2.04439***
	(0.2267)	(0.3737)
<i>FI*Distance=2</i>	-0.32969***	-0.94937***
	(0.0972)	(0.1547)
<i>FI*Distance=3</i>	-0.08525**	-0.04222
	(0.0425)	(0.0714)
<i>FI*Distance=4</i>	-0.01775	-0.04016
	(0.0243)	(0.0423)
<i>FI*Distance=5</i>	-0.00178	0.04232
	(0.0190)	(0.0331)
<i>FA*Distance=1</i>	0.15340	0.16033
	(0.1918)	(0.3648)
<i>FA*Distance=2</i>	-0.37640***	-0.49465***
	(0.0757)	(0.1517)
<i>FA*Distance=3</i>	-0.15627***	0.02975
	(0.0326)	(0.0677)
<i>FA*Distance=4</i>	-0.03998**	-0.01268
	(0.0184)	(0.0408)
<i>FA*Distance=5</i>	-0.01935	-0.00135
	(0.0139)	(0.0322)
<i>FA*FI</i>	0.01102	0.00411
	(0.0190)	(0.0334)
<i>FA*FI*Distance=1</i>	0.72679*	1.55643**
	(0.4305)	(0.7526)
<i>FA*FI*Distance=2</i>	0.42763**	1.18468***
	(0.1812)	(0.2979)
<i>FA*FI*Distance=3</i>	0.20698***	-0.08082
	(0.0793)	(0.1361)
<i>FA*FI*Distance=4</i>	0.03279	-0.00725
	(0.0439)	(0.0780)
<i>FA*FI*Distance=5</i>	-0.00626	-0.01874
	(0.0345)	(0.0614)
<i>N</i>	1,209,343	888,637
<i>R squared</i>	0.140	0.170

Notes: Column (1) defines a new idea using the text of titles only. Column (2) defines a new idea using the Unified Medical Language Systems® (UMLS®), which identifies terms that are synonyms. The outcome variable is a binary variable for adoption within 5 years after the birth of a new idea. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects, and the rank of the journal in which the idea appeared interacted with network distances.

References

1. Abramo, G., C.A. D' Angelo, and G. Murgia, *Gender differences in research collaboration*. Journal of Informetrics, 2013. **7**(4): p. 811-822.
2. Alcácer, J. and M. Gittelman, *Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations*. The Review of Economics and Statistics, 2006. **88**(4): p. 774-779.
3. Almeida, P. and B. Kogut, *Localization of Knowledge and the Mobility of Engineers in Regional Networks*. Management Science, 1999. **45**(7): p. 905-917.
4. Association of American Medical Colleges, *The Diversity Research Forum: Getting to Institutional Excellence: Ensuring the Integration of Diversity in Academic Medicine*. 2009: Boston, Massachusetts.
5. Azoulay, P., J. Graff Zivin, and J. Wang, *Superstar extinction*. Quarterly Journal of Economics, 2010. **25**: p. 549-589.
6. Baker, J.G., *Gender, race and Ph.D. completion in natural science and engineering*. Economics of Education Review, 1998. **17**(2): p. 179-188.
7. Beaudry, C. and V. Larivière, *Which gender gap? Factors affecting researchers' scientific impact in science and medicine*. Research Policy, 2016. **45**(9): p. 1790-1817.
8. Borjas, G.J. and K.B. Doran, *The Collapse of the Soviet Union and the Productivity of American Mathematicians**. The Quarterly Journal of Economics, 2012. **127**(3): p. 1143-1203.
9. Borjas, G.J. and K.B. Doran, *How high-skill immigration affects science: Evidence from the collapse of the USSR*. Innovation Policy and the Economy, 2015. **15**(1): p. 1-25.
10. Buffington, C., et al., *STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census*. American Economic Review, 2016. **106**(5): p. 333-38.
11. Card, D., et al., *Are referees and editors in economics gender neutral?* The Quarterly Journal of Economics, 2020. **135**(1): p. 269-327.
12. Catalini, C., *Microgeography and the Direction of Inventive Activity*. 2017, Rotman School of Management Working Paper No. 2126890.
13. Ceci, S.J., et al., *Women in academic science: A changing landscape*. Psychological science in the public interest, 2014. **15**(3): p. 75-141.
14. Cikara, M., L. Rudman, and S. Fiske, *Dearth by a Thousand Cuts? Accounting for Gender Differences in Top-Ranked Publication Rates in Social Psychology*. The Journal of social issues, 2012.

68(2): p. 263-285.

15. Collins, H., *Gravity's Shadow. The Search for Gravitational Waves*. 2004, Chicago: University of Chicago Press.
16. Correll, S.J., *Constraints into preferences: Gender, status, and emerging career aspirations*. American sociological review, 2004. **69**(1): p. 93-113.
17. Ding, W.W., F. Murray, and T.E. Stuart, *Gender Differences in Patenting in the Academic Life Sciences*. Science, 2006. **313**(5787): p. 665-667.
18. Duch, J., et al., *The Possible Role of Resource Requirements and Academic Career-Choice Risk on Gender Differences in Publication Rate and Impact*. PLOS ONE, 2012. **7**(12): p. e51332.
19. Dupas, P., et al., *Gender and the Dynamics of Economics Seminars*. National Bureau of Economic Research Working Paper Series, 2021. **No. 28494**.
20. Fong, E.A. and A.W. Wilhite, *Authorship and citation manipulation in academic research*. PLOS ONE, 2017. **12**(12): p. e0187394.
21. Freeman, R.B. and W. Huang, *Collaboration: Strength in diversity*. Nature, 2014. **513**(7518): p. 305-305.
22. Gardiner, M., et al., *Show me the money! An empirical analysis of mentoring outcomes for women in academia*. Higher Education Research & Development, 2007. **26**(4): p. 425-442.
23. Ghiasi, G., V. Larivière, and C.R. Sugimoto, *On the Compliance of Women Engineers with a Gendered Scientific System*. PLOS ONE, 2016. **10**(12): p. e0145931.
24. Gilbert, N.G., *Referencing as persuasion*. Social studies of science, 1977. **7**(1): p. 113-122.
25. Ginther, D.K. and S. Kahn, *Does science promote women? Evidence from academia 1973-2001*, in *Science and engineering careers in the United States: An analysis of markets and employment*. 2009, University of Chicago Press. p. 163-194.
26. Ginther, D.K., et al., *Race, ethnicity, and NIH research awards*. Science, 2011. **333**(6045): p. 1015-1019.
27. Glaeser, E.L., et al., *Growth in Cities*. Journal of Political Economy, 1992. **100**(6): p. 1126-1152.
28. Halpern, D.F., *Sex differences in cognitive abilities*. 2000: Psychology press.
29. Halpern, D.F., et al., *The Science of Sex Differences in Science and Mathematics*. Psychological science in the public interest : a journal of the American Psychological Society, 2007. **8**(1): p. 1-51.
30. Hebert, Camille. *Gender Stereotypes and Entrepreneur Financing*. 2020, Working Paper.
31. Hofstra, B., et al., *The Diversity–Innovation Paradox in Science*. Proceedings of the National Academy of Sciences, 2020. **117**(17): p. 9284.
32. Jacobs, J., *The Death and Life of Great American Cities*. 1961: New York: Vantage Books.
33. Jaffe, A.B., M. Trajtenberg, and R. Henderson, *Geographic localization of knowledge spillovers as evidenced by patent citations*. the Quarterly journal of Economics, 1993. **108**(3): p. 577-598.
34. Kachchaf, R., et al., *Career–life balance for women of color: Experiences in science and engineering academia*. Journal of Diversity in Higher Education, 2015. **8**(3): p. 175.
35. Kaiser, D., *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics*. 2005: Chicago: University of Chicago Press.
36. Knobloch-Westerwick, S., C.J. Glynn, and M. Huge, *The Matilda effect in science communication:*

an experiment on gender bias in publication quality perceptions and collaboration interest. Science communication, 2013. **35**(5): p. 603-625.

37. Koffi, M., *Innovative ideas and gender inequality.* 2021, Working Paper.

38. Krugman, P., *Increasing Returns and Economic Geography.* Journal of Political Economy, 1991. **99**(3): p. 483-499.

39. Larivière, V., et al., *Bibliometrics: Global gender disparities in science.* Nature, 2013. **504**(7479): p. 211-213.

40. Larivière, V., et al., *Sex differences in research funding, productivity and impact: an analysis of Québec university professors.* Scientometrics, 2011. **87**(3): p. 483-498.

41. Lerchenmüller, C., M.J. Lerchenmueller, and O. Sorenson, *Long-term analysis of sex differences in prestigious authorships in cardiovascular research supported by the National Institutes of Health.* Circulation, 2018. **137**(8): p. 880-882.

42. Ley, T.J. and B.H. Hamilton, *The Gender Gap in NIH Grant Applications.* Science, 2008. **322**(5907): p. 1472-1474.

43. Lincoln, A.E., et al., *The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s.* Social Studies of Science, 2012. **42**(2): p. 307-320.

44. Marschke, G., et al., *Last Place? The Intersection of Ethnicity, Gender, and Race in Biomedical.* AEA papers and proceedings. American Economic Association, 2018. **108**(5): p. 222-227.

45. Merton, R.K., *The Matthew effect in science: The reward and communication systems of science are considered.* Science, 1968. **159**(3810): p. 56-63.

46. Merton, R.K., *The sociology of science: Theoretical and empirical investigations.* 1973: University of Chicago Press.

47. Moed, H.F., *Citation analysis in research evaluation.* 2005: Springer Netherlands.

48. Moed, H.F. and E. Garfield, *In basic science the percentage of "authoritative" references decreases as bibliographies become shorter.* Scientometrics, 2004. **60**(3): p. 295-303.

49. Moed, H.F. and T.N. Van Leeuwen, *Impact factors can mislead.* Nature, 1996. **381**(6579): p. 186-186.

50. Oh, S.S., et al., *Diversity in Clinical and Biomedical Research: A Promise Yet to Be Fulfilled.* PLOS Medicine, 2015. **12**(12): p. e1001918.

51. Packalen, M. and J. Bhattacharya, *Age and the Trying Out of New Ideas.* Journal of Human Capital, 2019. **13**(2): p. 341-373.

52. Packalen, M. and J. Bhattacharya, *NIH funding and the pursuit of edge science.* Proceedings of the National Academy of Sciences, 2020, 117(22): 12011-12016.

53. Peri, G., *Determinants of knowledge flows and their effect on innovation.* Review of Economics and Statistics, 2005. **87**(2): p. 308-322.

54. Reagans, R. and E.W. Zuckerman, *Networks, diversity, and productivity: The social capital of corporate R&D teams.* Organization science, 2001. **12**(4): p. 502-517.

55. Rhoten, D., E. O'Connor, and E.J. Hackett, *The Act of Collaborative Creation and the Art of Integrative Creativity: Originality, Disciplinarity and Interdisciplinarity.* Thesis Eleven, 2009. **96**(1): p. 83-108.

56. Romer, P.M., *Endogenous technological change.* Journal of political Economy, 1990. **98**(5, Part

2): p. S71-S102.

57. Ross, M.B.B.G.R.M.-G.E.B.B.A.W.a.J.I.L., *Rosalind Franklin at Scale: Credit and Women in Science*. 2021.

58. Rossiter, M.W., *The Matthew Matilda Effect in Science*. Social Studies of Science, 1993. **23**(2): p. 325-341.

59. Sarsons, H., et al., *Gender differences in recognition for group work*. Journal of Political Economy, 2021. **129**(1): p. 000-000.

60. Singh, J., *Collaborative Networks as Determinants of Knowledge Diffusion Patterns*. Management Science, 2005. **51**(5): p. 756-770.

61. Staudt, J., et al., *High-impact and transformative science (HITS) metrics: Definition, exemplification, and comparison*. PLOS ONE, 2018. **13**(7): p. e0200597.

62. Thompson, P. and M. Fox-Kean, *Patent Citations and the Geography of Knowledge Spillovers: A Reassessment*. American Economic Review, 2005. **95**(1): p. 450-460.

63. Torvik, V.I. and S. Agarwal, *Ethnea--an instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database*, in *International Symposium on Science of Science*. 2016: Library of Congress, Washington DC.

64. Torvik, V.I. and N.R. Smalheiser, *Author name disambiguation in MEDLINE*. ACM Transactions on Knowledge Discovery from Data (TKDD), 2009. **3**(3): p. 1-29.

65. Torvik, V.I., et al., *A probabilistic similarity metric for Medline records: A model for author name disambiguation*. Journal of the American Society for information science and technology, 2005. **56**(2): p. 140-158.

66. Uzzi, B., et al., *Atypical Combinations and Scientific Impact*. Science, 2013. **342**(6157): p. 468-472.

67. Waldinger, F., *Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany*. Journal of Political Economy, 2010. **118**(4): p. 787-831.

68. Waldinger, F., *Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany*. The Review of Economic Studies, 2012. **79**(2): p. 838-861.

69. Wennerås, C. and A. Wold, *Nepotism and Sexism in Peer-Review*. Nature, 1997. **7**: p. 46-52.

70. West, J.D., et al., *The role of gender in scholarly authorship*. PloS one, 2013. **8**(7): p. e66212.

71. Woolgar, S., *Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy*. Science and Public Policy, 1991. **18**(5): p. 319-326.

72. Wuchty, S., B.F. Jones, and B. Uzzi, *The increasing dominance of teams in production of knowledge*. Science, 2007. **316**(5827): p. 1036-1039.

73. Zacchia, P., *Knowledge Spillovers through Networks of Scientists*. The Review of Economic Studies, 2019.

74. Zuckerman, H., *Scientific Elite*. 1977: New York: Free Press.

Appendix

Table A1. Stratify by the Gender Composition of Fields

	(1)	(2)
	Fields with fewer women	Fields with more women
<i>Distance=1</i>	2.49762**	3.07263**
	(0.05551)	(0.08089)
<i>Distance=2</i>	0.81257**	1.19272**
	(0.02431)	(0.03704)
<i>Distance=3</i>	0.27097**	0.25446**
	(0.01137)	(0.01705)
<i>Distance=4</i>	0.06086**	0.05233**
	(0.00682)	(0.00968)
<i>Distance=5</i>	0.01997**	0.00514
	(0.00511)	(0.00689)
<i>FI*Distance=1</i>	-0.65653**	-0.50167**
	(0.11973)	(0.15554)
<i>FI*Distance=2</i>	-0.08674*	-0.33791**
	(0.05232)	(0.07235)
<i>FI*Distance=3</i>	-0.06377**	-0.02228
	(0.02296)	(0.03289)
<i>FI*Distance=4</i>	-0.00158	0.02949
	(0.01297)	(0.01796)
<i>FI*Distance=5</i>	-0.00696	0.01447
	(0.00976)	(0.01272)
<i>FI</i>	0.02056*	-0.01021
	(0.01155)	(0.01147)
<i>FA*Distance=1</i>	-1.41401**	-1.49756**
	(0.17292)	(0.20417)
<i>FA*Distance=2</i>	-0.48011**	-0.63217**
	(0.07309)	(0.08978)

<i>FA*Distance=3</i>	-0.12976**	-0.09474**
	(0.03330)	(0.04166)
<i>FA*Distance=4</i>	-0.04205**	0.00699
	(0.01963)	(0.02326)
<i>FA*Distance=5</i>	-0.03730**	0.02324
	(0.01561)	(0.01748)
<i>FA*FI</i>	-0.01162	0.02081
	(0.01606)	(0.01543)
<i>FA*FI*Distance=1</i>	0.97202**	1.97184**
	(0.31841)	(0.36103)
<i>FA*FI*Distance=2</i>	-0.34525**	0.97238**
	(0.14261)	(0.16612)
<i>FA*FI*Distance=3</i>	0.02731	-0.13263*
	(0.06387)	(0.07672)
<i>FA*FI*Distance=4</i>	0.05313	0.00711
	(0.03644)	(0.04217)
<i>FA*FI*Distance=5</i>	0.04135	-0.00062
	(0.02916)	(0.03186)
<i>N</i>	3,689,320	3,680,184
<i>R squared</i>	0.07947	0.06520

Notes: This table reports the estimates used to generate Figure 6. We divide the full sample into two subsamples by the median of the share of women in the research field. Column (1) shows the estimation results for research fields with fewer women and Column (2) shows those for research fields with more women. The outcome variable is a binary variable for adoption within 5 years after a new idea is born. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects.

Table A2. Stratify by Journal Ranks

	(1)	(2)
	New ideas published in better	New ideas published in less

	journals	good journals
<i>Distance=1</i>	5.46148**	4.75321**
	(0.05608)	(0.13714)
<i>Distance=2</i>	1.43548**	0.86863**
	(0.02466)	(0.06495)
<i>Distance=3</i>	0.38728**	0.30293**
	(0.01139)	(0.03059)
<i>Distance=4</i>	0.07182**	0.12325**
	(0.00674)	(0.01714)
<i>Distance=5</i>	0.01260**	0.02817**
	(0.00500)	(0.01197)
<i>FI*Distance=1</i>	-1.56407**	-0.32495
	(0.11476)	(0.27120)
<i>FI*Distance=2</i>	-0.41102**	0.04739
	(0.05090)	(0.13041)
<i>FI*Distance=3</i>	-0.08141**	-0.19165**
	(0.02248)	(0.06052)
<i>FI*Distance=4</i>	0.01910	-0.06511**
	(0.01261)	(0.03253)
<i>FI*Distance=5</i>	0.00816	-0.01297
	(0.00935)	(0.02252)
<i>FI</i>	0.00615	-0.01205
	(0.01016)	(0.01913)
<i>FA*Distance=1</i>	-2.23726**	-0.35294
	(0.16364)	(0.33860)
<i>FA*Distance=2</i>	-0.59558**	-0.08001
	(0.06898)	(0.15800)
<i>FA*Distance=3</i>	-0.19859**	0.07878
	(0.03123)	(0.07606)
<i>FA*Distance=4</i>	-0.02942	-0.00485
	(0.01802)	(0.04225)

<i>FA*Distance=5</i>	-0.00740	0.00483
	(0.01417)	(0.03058)
<i>FA*FI</i>	-0.00889	0.02563
	(0.01372)	(0.02631)
<i>FA*FI*Distance=1</i>	3.08651**	1.75184**
	(0.28867)	(0.63815)
<i>FA*FI*Distance=2</i>	0.74070**	0.01723
	(0.12942)	(0.30454)
<i>FA*FI*Distance=3</i>	-0.01897	-0.01020
	(0.05776)	(0.14914)
<i>FA*FI*Distance=4</i>	0.05806*	-0.03705
	(0.03250)	(0.08090)
<i>FA*FI*Distance=5</i>	0.01361	0.02161
	(0.02578)	(0.05780)
<i>N</i>	5,640,252	1,742,389
<i>R squared</i>	0.06293	0.08794

Notes: This table reports the estimates used to generate Figure 7. Column (1) shows the estimation results for new ideas which are first published in journals that ranked in the top quartile, and Column (2) shows the estimates for the rest of the new important ideas in our sample. The outcome variable is a binary variable for adoption within 5 years after a new idea is born. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects.

Table A3. New ideas introduced by single teams

	Adoption	
	(1)	(2)
<i>Distance=1</i>	1.36289***	0.93982***
	(0.0621)	(0.0782)
<i>Distance=2</i>	0.56577***	0.53408***
	(0.0290)	(0.0365)

<i>Distance=3</i>	0.14496***	0.15608***
	(0.0132)	(0.0165)
<i>Distance=4</i>	0.04829***	0.05279***
	(0.0073)	(0.0092)
<i>Distance=5</i>	0.00904*	0.00811
	(0.0051)	(0.0064)
<i>FI*Distance=1</i>	1.27531***	0.89744***
	(0.1297)	(0.1337)
<i>FI*Distance=2</i>	-0.20633***	-0.22179***
	(0.0609)	(0.0626)
<i>FI*Distance=3</i>	-0.05107*	-0.06501**
	(0.0268)	(0.0275)
<i>FI*Distance=4</i>	0.00812	0.01366
	(0.0141)	(0.0145)
<i>FI*Distance=5</i>	0.00416	0.00290
	(0.0097)	(0.0100)
<i>FI</i>	-0.00763	-0.01022
	(0.0102)	(0.0107)
<i>FA*Distance=1</i>	-0.57294***	-0.72899***
	(0.1801)	(0.1871)
<i>FA*Distance=2</i>	-0.24735***	-0.37049***
	(0.0826)	(0.0866)
<i>FA*Distance=3</i>	0.02394	0.05662
	(0.0354)	(0.0371)
<i>FA*Distance=4</i>	0.00697	0.01311
	(0.0191)	(0.0201)
<i>FA*Distance=5</i>	-0.00806	-0.00395
	(0.0139)	(0.0146)
<i>FA*FI</i>	0.02485**	0.02908**
	(0.0119)	(0.0123)
<i>FA*FI*Distance=1</i>	0.92872***	0.59653*

	(0.3241)	(0.3406)
<i>FA*FI*Distance=2</i>	0.21823	0.41126***
	(0.1516)	(0.1591)
<i>FA*FI*Distance=3</i>	-0.18700***	-0.21365***
	(0.0664)	(0.0700)
<i>FA*FI*Distance=4</i>	0.02133	0.01278
	(0.0348)	(0.0367)
<i>FA*FI*Distance=5</i>	0.00519	0.00044
	(0.0254)	(0.0266)
<i>Journal Rank * distances FEs</i>	No	Yes
<i>N</i>	3,125,711	2,947,851
<i>R squared</i>	0.077	0.081

Notes: The sample consists of observations involving new ideas introduced by a single team. The outcome variable is a binary variable for adoption within 5 years after a new idea is born. FI and MI indicate the gender of the closest innovator. FA and MA indicate the gender of a potential adopter. Journal rank is based on the Scimago journal rank index of the journal that first published the new idea. All estimated coefficients are multiplied by 100. All regressions control for individual fixed effects and idea fixed effects.