PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

A scientist like me: demographic analysis of biology textbooks reveals both progress and long-term lags

Sara Wood[1], Jeremiah A. Henning[2], Luoying Chen[1], Taylor McKibben[1], Michael L. Smith[3,4,5], Marjorie Weber[6], Ash Zemenick[6] and Cissy J. Ballen[1]

[1]Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA
[2]Department of Biology, University of South Alabama, Life Sciences, Mobile, AL 36688, USA
[3]Department of Collective Behavior, Max Planck Institute of Animal Behavior, Konstanz, Germany
[4]Centre for the Advanced Study of Collective Behaviour, and [5]Department of Biology, University of Konstanz, Konstanz, Germany
[6]Department of Plant Biology, Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA

JAH, 0000-0002-2214-4895; MLS, 0000-0002-3454-962X; MW, 0000-0001-8629-6284; AZ, 0000-0001-9889-5532; CJB, 0000-0002-4693-6117

Biological science practices

Cite this article: Wood S, Henning JA, Chen L, McKibben T, Smith ML, Weber M, Zemenick A, Ballen CJ. 2020 A scientist like me: demographic analysis of biology textbooks reveals both progress and long-term lags. *Proc. R. Soc. B* **287**: 20200877. http://dx.doi.org/10.1098/rspb.2020.0877

Received: 17 April 2020
Accepted: 29 May 2020

Subject Category:
Evolution

Subject Areas:
evolution

Keywords:
biology textbooks, representation, stem equity, role models, intersectionality

Author for correspondence:
Cissy J. Ballen
e-mail: mjb0100@auburn.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5018441.

THE ROYAL SOCIETY PUBLISHING

Textbooks shape teaching and learning in introductory biology and highlight scientists as potential role models who are responsible for significant discoveries. We explore a potential demographic mismatch between the scientists featured in textbooks and the students who use textbooks to learn core concepts in biology. We conducted a demographic analysis by extracting hundreds of human names from common biology textbooks and assessing the binary gender and race of featured scientists. We found that the most common scientists featured in textbooks are white men. However, women and scientists of colour are increasingly represented in contemporary scientific discoveries. In fact, the proportion of women highlighted in textbooks has increased in lockstep with the proportion of women in the field, indicating that textbooks are matching a changing demographic landscape. Despite these gains, the scientists portrayed in textbooks are not representative of their target audience—the student population. Overall, very few scientists of colour were highlighted, and projections suggest it could take multiple centuries at current rates before we reach inclusive representation. We call upon textbook publishers to expand upon the scientists they highlight to reflect the diverse population of learners in biology.

1. Introduction

Textbooks are one of the primary resources that undergraduate students use to learn science and are often required reading as part of coursework [1]. While conveying foundational concepts in a given discipline, textbooks highlight the historical work of influential scholars who have shaped the field. Whether intentionally or not, textbooks instil readers with ideas about who can contribute to science, technology, engineering or mathematical (STEM) fields [2]. Therefore, textbooks represent an important opportunity to shape students' existing stereotypes of who scientists are, have been and can be.

Student perceptions of who can do science influence their sense of belonging in STEM fields, which in turn affects their performance and retention [3]. Perceptions are shaped by environmental cues within a context, and previous work shows exposure to stereotypical representations of scientists impacts interest in science among women and students of colour [4–6]. Cheryan *et al.* [7] showed that women lost interest in computer science classrooms when objects from the room signalled that computer scientists are 'geeky' men (e.g. *Star Trek* posters). In this case, objects broadcasted stereotypes about a group, which discouraged people who did not fit that stereotype from pursuing that potential interest. Additionally, the lack of role models or visual representation of people of colour may lead to increased imposter syndrome among such groups. Imposter syndrome is the perception that one doesn't deserve their accomplishments, or a sense of intellectual phoniness [8,9]. Thus, without regular exposure to diverse, relatable role models, scientist stereotypes have the potential to be particularly harmful for students who identify with under-represented and/or marginalized groups. By contrast, exposing students to scientists from a diversity of backgrounds and identities has positive impacts on students' interest and achievement in STEM [10–17]. This impact can be long lasting: in one study, biology students exposed to examples of scientists from under-represented groups in class activities reported increased ability to personally relate to scientists up to six months later [18].

Despite concern about the impacts of frequent exposure to stereotypical representations of scientists, the extent to which scientists from diverse backgrounds and identities are included in undergraduate biology textbooks remains poorly understood. Previous work has explored textbook representation of women in six chapters across seven ecology textbooks from 2000 to 2005 [19]. They found that women were less represented than expected across all reasons for which they were cited (e.g. as a founder/innovator, working scientists, featured in pictures). Additionally, they found students who were provided materials that included women or scientists of colour throughout the semester were able to list more examples of them in an exercise at the end of the semester, demonstrating how modified course content can affect students' awareness of the participation of women in science.

Another study explored multiple axes of identity across three chapters of 12 geoscience textbooks using quantitative and qualitative approaches [20]. They found 94% of all founders/innovators, presented as contributing to major discoveries or innovations, to be men. A demographic analysis of photographs allowed authors to categorize race as Black, White or 'other'; of the twelve Black individuals shown in pictures, ten were in a single photograph used to demonstrate the problem of overpopulation. They concluded that geoscience education still reflects a scientific field of inquiry that is predominantly masculine and White.

Here, we fill critical holes in the literature by exploring the intersectional identities of scientists in introductory biology textbooks, and forecasting future representation compared to the student and general population, assuming current rates of change continue. An intersectionality perspective is critical as sociodemographic constructs such as race and gender also interact with one another and with other social categories (e.g. socioeconomic status) to shape people's experiences [21]. For this reason, attaining a deep understanding of inequities in STEM fields requires consideration of the ways in which axes of identity intersect to create distinct identity configurations (e.g. [22]).

We chose to study introductory textbooks for several reasons. First, substantial evidence suggests introductory science courses are formative experiences for students who wish to pursue science. For example, performance in introductory STEM courses is an indicator for choosing STEM [23,24], and one of the most frequently cited reasons for leaving STEM is the challenging, overwhelming nature of introductory science courses [25]. Second, we were interested in texts that served a broad group of students, because almost all students must go through an introductory class before moving on to more specialized coursework.

Given the evidence that perceptions of scientists are important in our national efforts to promote inclusivity in classrooms [26], we sought to characterize the status of demographic representation of biologists across common contemporary biology textbooks in the United States, and how representation has changed over the history of biology research. We addressed the following specific questions. (i) In contemporary textbooks, does the demographic makeup of scientists represented change over the history of biological discovery? (ii) Are the proportion of women scientists featured in biology textbooks representative of the makeup of active biologists at the time of discovery? (iii) What is the overall demographic (binary gender, race) representation of scientists in biology textbooks, and how does this compare to the makeup of the student population?

2. Material and methods

(a) Textbook selection

We explored seven commonly used biology textbooks in introductory biology classes across the United States. We identified these texts using methods described in [27], which we will summarize here. Researchers identified the single largest United States four-year university in each state using information from the US Department of Education (CollegeStats.org). Then, using each university website, they identified the ten most frequently used introductory biology textbooks assigned to students on a biological sciences track, focusing on the most recent edition of the texts at the time of the study (ranging from 2016 to 2019). Because our methods required electronic versions of each textbook, our final list represents a convenience sample of textbooks that were electronically available ($n = 7$).

(b) Identification of scientists and demographic assignments

We extracted the names of all scientists listed in the indices of each textbook using Python (Python Software Foundation; full methods in electronic supplementary material). From the seven textbooks, we identified 1151 names in the indices. After removing the names of non-scientists and unverifiable entries, there were 1107 scientists (average of 164 scientists per textbook). In some cases, an individual scientist was highlighted in more than one textbook, and so they are represented multiple times in our total list. For example, Carolus Linnaeus, Charles Darwin and Gregor Mendel were mentioned in all seven textbooks; Hopi Hoekestra, Jane Goodall and Rosemary Grant were mentioned in three out of the seven textbooks. We decided to include these names as a proxy for actual exposure students have to scientists across these texts (electronic supplementary material, table S1). However, note that our results and conclusions are the same with or without the replicate scientists. Additionally, 14 out of the 1107 scientists' race was unknown (table 1), but this is unlikely to impact our results. For each individual, we recorded their binary gender, race and year of published work. For binary gender, scientists were identified as either men or women based on the pronouns used in the textbooks. This assumes that all scientists represented were cisgender and identified with gender that aligns with their gender presentation. Unfortunately, data on self-reported gender identity that is inclusive to cisgender, transgender, non-binary and/or gender-nonconforming people was not available. If the gender could not be inferred from the textbook, Wikipedia profile information was used. In addition, some scientists have dedicated web pages for their research, which occasionally included demographic information such as self-identified gender, race and sometimes birth

Table 1. Binary gender, racial and year of publication differences in the citations of scientists in seven commonly used biology textbooks.

factor	χ^2	d.f.	*p*-value
binary gender	270.03	1	<0.0001
race	1385.13	3	<0.0001
year of publication	4.71	10	0.91
binary gender × race	765.44	3	<0.0001
race × year of publication	15.23	30	0.99
binary gender × year of publication	33.74	10	0.0002
race × binary gender × year of publication	106.03	30	<0.0001

year. If available, we cross-referenced textbook or Wikipedia information with self-reported information. For racial assignments, we followed the National Institutes of Health guidelines for defining racial categories in the context of the United States [28]: American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, or White. These are based on standards for the classification of federal data on race, commonly used for federal data collection purposes including the decennial census. In instances when the race was not reported, or the race reported did not fit into the categories laid out by these guidelines, that scientist's race was labelled as unknown. For the purposes of this research we use 'scientists of colour' to describe all scientists who are not White, while acknowledging this does not recognize the variation within and among groups. Some individuals in these groups do not identify with this term in a singular way, and some reject this term altogether. However, we chose to use it over 'non-White' because we reject the idea of positioning whiteness as the default. We recognize these categories can be problematic because they are defined by an authority; they do not leave room for or recognize people who identify as mixed race and are limiting because textbooks draw from an international pool of scientists. We further recognize that binary gender and race are only two of many human social identities that have subpopulations which are marginalized and under-represented in STEM fields; while imperfect, our categories allow us to establish baselines of identity representation in the most commonly used biology textbooks in the US.

Beyond the absolute representation of scientists in textbooks, we were also interested in quantifying whether time of publication played a role in determining the ratio of gender and race identities portrayed in textbooks. We recorded the year of published work for each study highlighted, defining the year of published work as the year the work highlighted (i.e. cited) in the textbook was published in the scientific literature. We predicted that time of discovery and representation of scientists of colour and women would positively covary.

(c) Comparison to academic population

To determine whether the proportion of a demographic group of scientists featured in biology textbooks was the same as one would expect based on their abundance as active biologists, we developed a proxy measurement for the proportion of men and women biologists in academia over the last 50 years. We used the National Science Board's *Science and Engineering Indicators*, which measured the approximate number of men and women tenured professors in life sciences from 1973 to 2010. We chose to investigate tenured professors because we reasoned that by the time researchers are considered eminent players in their field, and most eligible to be highlighted in a textbook, they are likely tenured. In order to test this assumption, we needed to determine when the scientists published the work that was featured in textbooks, and whether their inclusion in a textbook occurred after tenure. Previous research shows that the majority of professors attain tenure between the ages of 40 and 44 [29,30]. Because we knew the dates of publication for each textbook, we could subtract the scientists' year of birth from this date if we could identify their year of birth. In order to identify scientists' date of birth, we scraped Wikipedia pages by obtaining the XML source through MediaWiki API (en.wikipedia.org/w/api.php), looking for textbook authors with information available on Wikipedia, and extracting birth year from the source code (see electronic supplementary materials for more information). Through this approach, we obtained the birth year of 355 scientists. By subtracting the birth year from the year of publication of work cited in the textbooks, we found the average age of scientists at the time of their notable discovery was 45 years old (15.2 s.d.). Therefore, most scientists whose work is published in textbooks have probably already achieved tenure.

Due to availability of data, we focus here on men and women life scientists; we acknowledge that future investigations on race/ethnicity over time would be of interest to the scientific community. Unfortunately, we could not address whether textbook representation based on race matched representation of life scientists with any meaningful level of resolution, as we are able to with binary gender. Data on race was only available for total faculty over time (i.e. available data was not broken up by scientist rank), and thus low sample sizes prohibited our ability to draw meaningful conclusions.

(d) Data analysis

To analyse the demographic representation of scientists over the history of biological discovery, we used descriptive statistics and linear mixed effects model using the *lme4* package in R version 3.6.0 [31,32]. Binary gender, race, year of publication as well as all possible interaction terms between binary gender, race and year of publication were set as fixed effects, and textbook was included as a random effect. Using this analysis, we were also able to examine overall differences in demographic representation of scientists in biology textbooks. To account for differences in the absolute number of citations included in each textbook and the different sampling time periods, we converted counts to proportions. We determined significant main effects and interaction effects via Wald tests using the *car* package [33].

To understand if the change in representation of women scientists within textbooks reflects the rate of active, tenured professors who were women at the time of discovery, we performed a chi-square goodness of fit test to measure deviations between observed textbook citations and the expected number of citations by men and women scientists per 10-year period, assuming the rate of citations would be proportional to the gender ratios of scientists at the time.

We calculated the predicted number of citations for a given year by multiplying the proportion of men and women scientists by the number of total scientists cited in the textbooks for a one year time period—for example in 1973: 0.93 men × 12 scientists = 11.15 men scientists; 0.071 × 12 scientists = 0.85 women scientists. Next, we determined the 'observed' textbook citations by summing the total number of men and women scientists for each year cited within all the textbooks. Next, we conducted goodness of fit tests after combining measures across decades (1973–1979, 1980–1989, 1990–1999, 2000–2010) to maximize the number of replicates compared (table 2). We removed 1979, 1983, 1987 and 1990 from our analysis because we had less than five overall textbook citations for those given years. We also conducted goodness of fit tests on each year individually (electronic supplementary material, table S2), and conducted a power analysis to determine our ability to detect deviations between observed and expected values (electronic supplementary material, figures S1 and S2).

Table 2. Deviation in observed number of textbook reference scientists compared to the background rate of life scientists who are women across the 1970s, 1980s, 1990s, 2000s and all years combined.

decade	χ^2	d.f.	*p*-value
1970s	6.597	5	0.252
1980s	6.791	8	0.559
1990s	8.297	8	0.405
2000s	13.034	10	0.222
all years	45.977	34	0.082

As a final exercise, we were interested in extrapolating our results to determine how long it would take for the representation of gender and race in textbooks to reflect (i) the population of the United States and (ii) the population of students graduating with degrees in biological sciences from postsecondary institutions. First, we modelled the proportion of scientists from gender and racial groups as a function of year using linear, quadratic and cubic functions. Next, we determined the 'best fit' relationship by looking at the coefficient of determination. Equations from best fit models for each gender and racial group can be found in electronic supplementary material, table S3. From our best fit model, we extrapolated this relationship into the future to understand how group representation in textbooks will change through time, assuming that past demographic shifts will continue at the same rate. While this is a large assumption, it reflects an estimation of demographic shifts assuming status quo changes into the future. We acknowledge that these are relatively simple estimates, but our aim was to understand the magnitude of timeframes to reach textbook representation that matches the general and student population.

3. Results

(a) In contemporary textbooks, does the demographic makeup of scientists represented change over the history of biological discovery?

First, we ran descriptive statistics to examine the change in representation of binary gender and racial categories within textbook citations. We predicted that representation of women and other scientists of colour would increase in citations of more recent research due to the diversification of STEM disciplines over time. We observed that women scientists are represented more in contemporary citations than historical citations. For example, for research published between 1900 and 1999, contemporary textbooks featured 55 women scientists (approx. 10% of all highlighted scientists); for research published from 2000 to 2018, contemporary textbooks featured 87 women scientists (25% of all highlighted scientists). This could be due to greater recognition of scientists who are women or may represent the fact that the absolute number of scientists who are women has grown over time or both. Our linear mixed effects analysis echoed the descriptive statistics; we found citations of scientific literature published over recent decades included a higher proportion of women relative to prior publications (gender × year of publication: $\chi^2 = 33.74$, $p = 0.0002$; table 1).

Racial representation also shifted with year of publication of the work featured in the textbooks. For research published between 1900 and 1999, textbooks highlighted 19 scientists of colour (3% of all highlighted scientists). For research published between 2000 and 2018, however, textbooks highlighted 27 scientists of colour (8% of all highlighted scientists). Results from the linear mixed model did not show a significant shift in the citations from scientists of colour when comparing across the years of publication (race × year of publication: $\chi^2 = 15.23$, $p = 0.98$). However, we did observe a three-way interaction between race, gender and year of publication ($\chi^2 = 106.0$, $p < 0.0001$) which indicates that the representation of certain groups increased through time (White women and Asian men), while others decreased (White men) and representation of some groups (Asian women, Black women, Hispanic men and women) do not significantly change over time. We observed significant underrepresentation (Asian and Hispanic women) or no representation (Black women) of women scientists of colour (figure 1). We observed some variation among textbooks with respect to representation over time (electronic supplementary material, figure S3).

(b) Are the proportion of women scientists featured in biology textbooks representative of the makeup of active biologists at the time of discovery?

We compared the representation of women scientists within textbooks to the abundance of women who were tenured biologists at the time of discovery by using chi-square goodness of fit tests to determine if our 'observed' scientific citations deviated from our 'predicted' scientific citations for each year. We found that citations of biologists who are women were remarkably proportional to the number of women biologists in the scientific workforce (figure 2, table 2).

(c) What is the overall demographic (binary gender, race) representation of scientists in biology textbooks, and how does this compare to the makeup of the general and student population?

When considering overall representation of scientists across textbooks, 145 scientists were women (13.1%) and 962 were men (86.9%), representing a 1 : 7 ratio of women to men ($\chi^2 = 270.1$, $p < 0.0001$). Only 6.67% of the scientists mentioned across textbooks were scientists of colour ($\chi^2 = 1385.1$, $p < 0.0001$; figure 2). These values do not reflect the demographic makeup of the general population or biology student population in the United States (table 3), and we questioned how long until women and scientists of colour are represented in biology textbooks at the same proportions as they are represented in the general and student population.

(d) How long until demographic representation in biology textbooks reflects that of the United States population and undergraduate biology student population?

The shift towards more inclusive representation within biology textbooks is occurring at different rates among binary gender and racial groups. Assuming that observed demographic shifts in textbook citations will continue at the

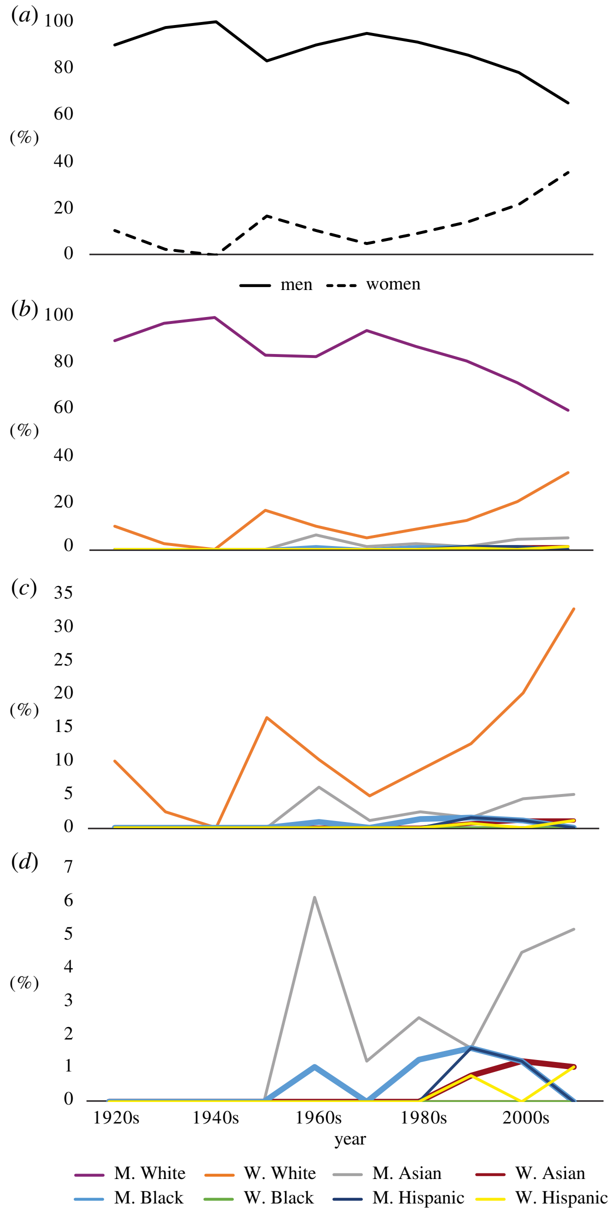

Figure 1. Demographic representation of scientists featured in biology textbooks, shown as proportions over time of scientific discovery. (*a*) Representation of scientists in terms of binary gender. (*b*) Representation of scientists in terms of binary gender and race. The majority of highlighted scientists were White men (purple line) and White women (orange line). (*c*,*d*) We increased resolution of the data by removing White men and then White women, respectively; (*d*) shows the values of scientists of colour featured (grey line: Asian men; red line: Asian women; light blue line: Black men; green line: Black women; navy blue line: Hispanic men; yellow line: Hispanic women). (Online version in colour.)

same rate, we estimated a best fit line to predict how group representation will change over time. Based on our extrapolations, we predicted that women representation in textbooks will reflect the general population (49%) in about 10 years, but will not reflect biology students (60%) for another 18 years (figure 3). Extrapolations of our textbook citation data of White scientists show they will decline to reach representation of the general population in approximately 55 years and the student population in 90 years. Our data show that Asian scientists in textbooks currently reflect Asian populations within the United States (6%), but will not reflect the biology student makeup (15.2%) for approximately 50 years (figure 4). Some of our estimates, however, revealed more

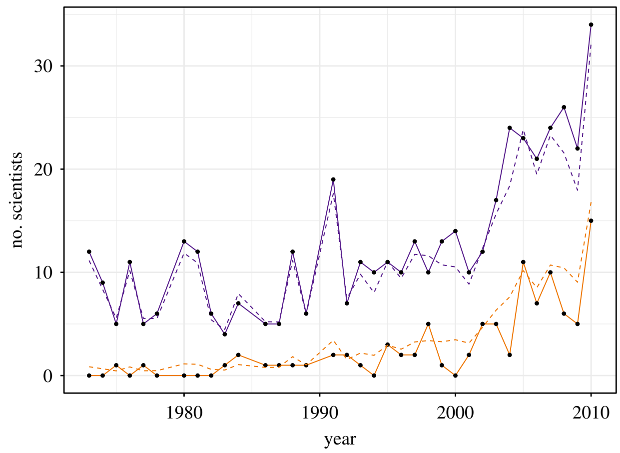

Figure 2. Observed (solid line) textbook citations of women (orange) and men (purple) relative to the predicted number of scientists (dashed line) based on the representation of binary gender across time. (Online version in colour.)

pessimistic projections: if textbook citations from Black/African American scientists continue at the same rate, it will take over 1000 years to reflect the general population in the United States (14%), and nearly 500 years to reflect the biology student population (7.7%). Among Hispanic/Latinx scientists highlighted in textbooks, we project 45 years until they reflect general public representation (16%) and 30 years until they reflect student populations (11.3%; figure 4).

4. Discussion

By examining scientists highlighted across seven common contemporary biology textbooks in the United States, we report on the changes in representation of binary gender and racial categories over time. We found higher representation of contemporary women and scientists of colour compared to historical researchers. Results from our linear mixed model showed that representation of some specific demographic subsets increased through time (White women and Asian men), while others remained significantly under-represented (Asian & Hispanic women) or not represented (Black women). Using an intersectional approach allowed us to understand trends at a higher resolution than considering only binary gender or race alone. To our knowledge, this is the first study to investigate biology textbooks through an intersectionality lens at this scale.

For binary gender, but not race, we were able to assess whether textbook representation of women aligned with the representation of tenured life science faculty who were women and found strong alignment between the observed and expected proportions. This suggests that with respect to binary gender representation, biology textbooks are accurately reflecting the demographic composition of biologists and the changes in the past 100 years. This finding contrasts similar studies in other fields that showed the underrepresentation of women in textbooks relative to their contributions to published scientific work (e.g. ecology textbooks [19]; geology textbooks [20]).

Taken cumulatively, however, there is an underrepresentation of relatable role models for students who are women or students of colour. For example, while over half of the United States population are women (US Census 2010), and

Table 3. Racial and binary gender profile of scientists represented in textbooks compared with that of the United States population and population of biology students.

race category	representation text (%)	general population[a] (%)	biology student population[b] (%)
American Indian/Alaskan native	0.0	2.0	0.4
Asian	2.9	6.0	15.2
Black/African American	0.6	14.0	7.7
Latinx/Hispanic	0.6	16.0	11.3
native Hawaiian or other Pacific Islander	0.0	0.4	0.2
white	94.6	72.0	58.5
could not identify	1.3	N/A	3.9
binary gender			
men	86.9	49.2	40.0
women	13.1	50.8	60.0

[a]For data within the 2010 United States census, individuals sometimes fall into multiple racial groups [34].
[b]Bachelor's degrees in biological sciences conferred by postsecondary institutions, by race/ethnicity over 2015–2016 [35]. Here, students who fell into multiple racial groups were categorized as two or more races, and represented 3.7% of all students.

60% of students awarded biological sciences bachelor's degrees in the United States are women [35], biology textbooks highlight seven men for every one woman scientist. Additionally, while over 29% of the United States population are people of colour (table 3), as well as 35% of students awarded biological sciences bachelor's degrees in the United States [35], only 6.67% of the scientists mentioned in textbooks were people of colour. Some demographic groups, such as Black women, were not represented a single time across any of the textbooks we analysed. Finally, we forecasted when scientists from demographic groups will be equally represented in textbooks as they are in the general population and among biology students; while the interpretation from our projections has limitations, it shows a grim outlook for some under-represented scientists. For example, if Black authors continue to be featured in biology textbooks at the same rate, it will take over 1,000 years to reflect the general population in the United States, and nearly 500 years to reflect the biology student population.

While demographics in the United States continue to diversify [36], a demographic mismatch between 'who students aspire to be' and 'who currently occupies science professions' intensifies. Many people have at least one axis of their identity which is negatively stereotyped, marginalized or under-represented [37], and the presence of role models is critical to intellectual growth and development [38]. Role models are inspiring [39] and can increase retention among undergraduates pursuing science degrees [40]. Role models are particularly important for under-represented groups who may not otherwise have access to mentors that share salient elements of their identity [41]. For example, many women are discouraged because they don't think a career in science is compatible with having a family [42]. However, in the presence of relatable, high-achieving women, students perform better and report higher self-esteem and science self-efficacy [14,39,43]. Similarly, previous research documents students of colour losing interest in science because they perceive that it lacks social value [41]. However, learning about the life and values of other scientists of colour increases interests in science and performance for such students [18]. It is also important to recognize that students possess multiple identities that interact, resulting in unique lived experiences. Black women, for example, encounter a combination of challenges that cannot be understood through their race or gender alone [22,44]. Efforts to expose undergraduate students to counter-stereotypical examples of scientists have the potential to narrow equity gaps and broaden participation of marginalized and under-represented groups in STEM.

To address the call to increase the diversity of scientist role models, classrooms have integrated counter-stereotypical examples of scientists in introductory biology using resources such as Scientist Spotlights (described in [16]) and Project Biodiversify (a repository of materials that provide examples from primary research and personal experiences from scientists that identify with under-represented groups in biology; www.projectbiodiversify.org). Some universities have also attempted to diversify the portraits depicting members of leadership (e.g. department chairs or deans) that decorate lobbies, conference rooms, hallways and lecture halls of universities—opting instead to highlight recent discoveries or research being conducted by graduate students [45]. While textbooks can be applauded for matching trends in the representation of women scientists, they are overall behind in representing the demographic composition of textbook consumers. We hope that this effort encourages textbook authors to diversify the scientists featured in biology textbooks, and we suggest one way to do this is to highlight contemporary research. In the meantime, we encourage educators to use alternative resources available that make classrooms more inclusive.

5. Limitations

The results of this research have limitations. For example, we scraped scientists' names from indices rather than the text body, and so we do not know the extent to which certain scientists are featured in the text (e.g. word counts), or how they are featured (e.g. with or without a photo). These are

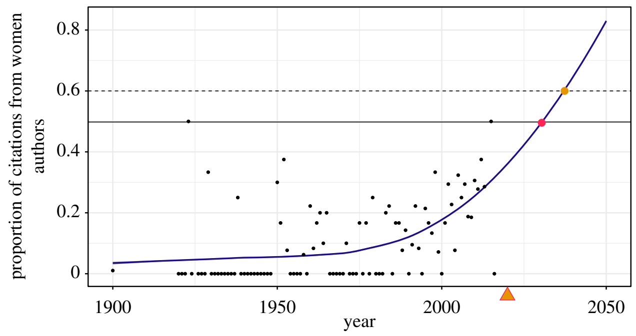

Figure 3. An extrapolation of women scientists as a function of year with best fit line, assuming that past demographic shifts in textbook representation will continue at the same rate in the future. According to this projection, relative to the present (orange triangle), women representation in textbooks will reflect the general population (49%; solid horizontal line) in approximately 10 years (red dot) and will reflect biology students (60%; dashed horizontal line) in 18 years (orange dot). (Online version in colour.)

important future directions for work in this area. Because our methods required electronic versions of each textbook, our final list of textbooks represented a convenience sample of textbooks that were electronically available. We did not include hardcover textbooks in the study and cannot rule out the possibility that this impacted our results. We also could not document aspects of publishing such as the decision process behind the selection of cited scientists. Additionally, in the current study, we used Wikipedia to acquire demographic information if the scientist did not have a professional website, but self-reported identities would have allowed us to expand beyond gender binaries, simplistic racial categories, and otherwise coarse and problematic delimitations of individual identity.

We accounted for a potential time lag when addressing whether textbooks over- or under-represented women over time by using the proportion of tenured professors as estimates of potential scientists to include in texts. However, this is a coarse estimate that might not capture whether the proportions of women featured in textbook matches those in the biology workforce who publish exemplary work. Finally, we reported demographic proportions of the biology student population as those who received bachelor's degrees in biological sciences. However, because women and students of colour are more likely to leave STEM fields, we probably underestimated the proportions of such students who used common textbooks in introductory biology. Thus, our forecast conclusions of time until different groups are equally represented in textbooks may be much longer than we can demonstrate at this point. Future work focusing on more nuanced and realistic axes of identity and more accurate proxies for student and scientist populations will strengthen this field of study.

6. Conclusion

Although comprehensive data on textbooks are scarce, our research shows that stereotypical scientists are still featured heavily in common biology textbooks. Future work will profit from an exploration of scientists as they are highlighted across disciplines, and as the workforce continues to diversify

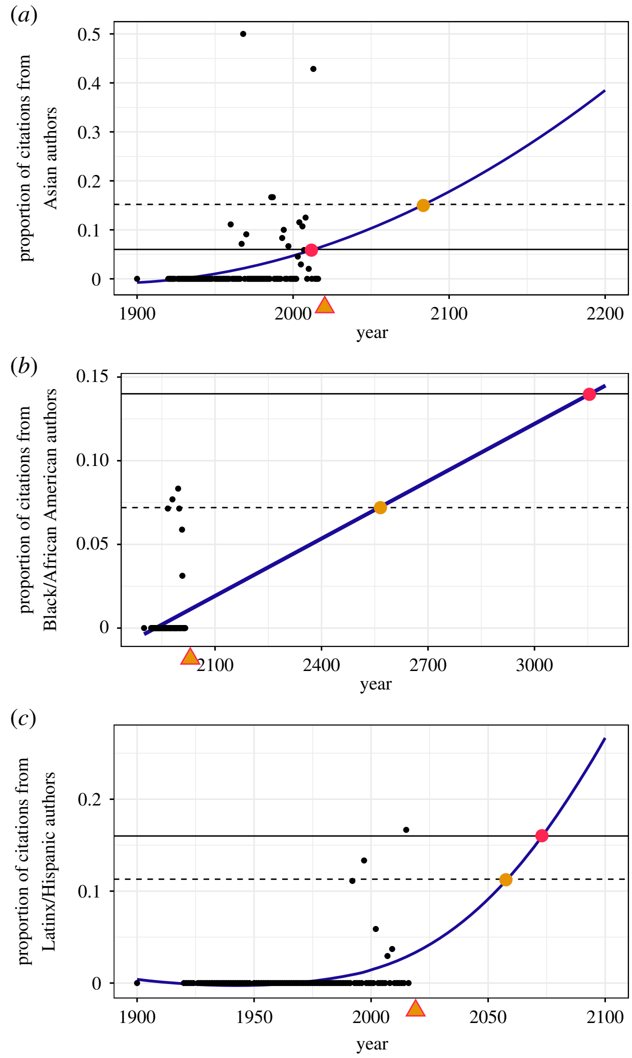

Figure 4. An extrapolation of Asian (*a*), Black/African American (*b*) and Latinx/Hispanic (*c*) scientists as a function of year with best fit lines. In these projections, trends in representation in textbooks are extrapolated from the current (orange triangle) to the future. Best fit lines estimate when textbook representation of these groups will reflect that of the general population (solid horizontal line, red dot) and the population of biology undergraduates (dashed horizontal line, orange dot). (Online version in colour.)

over time. We do not advocate for an erasure of the history of science, or intend to undermine the enormous contributions of individuals who laid the groundwork for contemporary biology. However, equally important in our efforts to communicate history is to show that science is a diverse enterprise and that anyone who is capable and interested in fundamental principles of life belongs in a science career.

Data accessibility. This article has no additional data.

Competing interests. We declare we have no competing interests.

Funding. This work was funded in part by grant nos. NSF DBI-1919462 (C.J.B.), NSF DEB-1831164 (M.W.) and NSF DBI-1708931 (A.Z.); DFG Centre of Excellence 2117 Centre for the Advanced Study of Collective Behaviour ID: 422037984 (M.L.S.).

Acknowledgements. We are grateful to the discipline-based education research group at Auburn University for valuable feedback on versions of the manuscript. MLS is a Simons Foundation Postdoctoral Fellow of the Life Sciences Research Foundation. Finally, we would like to thank four anonymous reviewers for their critical insights that greatly improved the manuscript.

References

1. Hilton J. 2016 Open educational resources and college textbook choices: a review of research on efficacy and perceptions. *Educ. Technol. Res. Dev.* **64**, 573–590. (doi:10.1007/s11423-016-9434-9)
2. Good JJ, Woodzicka JA, Wingfield LC. 2010 The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *J. Soc. Psychol.* **150**, 132–147. (doi:10.1080/00224540903366552)
3. Margolis J, Fisher A, Miller F. 2000 The anatomy of interest: women in undergraduate computer science. *Women's Stud. Q.* **28**, 104–127.
4. Cheryan S, Plaut VC, Handron C, Hudson L. 2013 The stereotypical computer scientist: gendered media representations as a barrier to inclusion for women. *Sex Roles.* **69**, 58–71. (doi:10.1007/s11199-013-0296-x)
5. DeWitt J, Archer L, Osborne J. 2013 Nerdy, brainy and normal: Children's and parents' constructions of those who are highly engaged with science. *Res. Sci. Educ.* **43**, 1455–1476. (doi:10.1007/s11165-012-9315-0)
6. Tanner KD. 2009 Learning to see inequity in science. *CBE—Life Sci. Educ.* **8**, 265–270.
7. Cheryan S, Plaut VC, Davies PG, Steele CM. 2009 Ambient belonging: how stereotypical cues impact gender participation in computer science. *J. Pers. Soc. Psychol.* **97**, 1045. (doi:10.1037/a0016239)
8. Stout JG, Dasgupta N, Hunsinger M, McManus MA. 2011 STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *J. Pers. Soc. Psychol.* **100**, 255–270. (doi:10.1037/a0021385)
9. Clance PR. 1985 *The impostor phenomenon: overcoming the fear that haunts your success*. Atlanta, GA: Peachtree Pub.
10. Yonas A, Sleeth M, Cotner S. 2020 In a 'scientist spotlight' intervention, diverse student identities matter. *J. Microbiol. Biol. Educ*. **21**. See https://www.asmscience.org/content/journal/jmbe/10.1128/jmbe.v21i1.2013.
11. Cheryan S, Master A, Meltzoff AN. 2015 Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Front. Psychol.* **6**, 1–8. (doi:10.3389/fpsyg.2015.00049)
12. Dee TS. 2004 Teachers, race, and student achievement in a randomized experiment. *Rev. Econ. Stat.* **86**, 195–210. (doi:10.1162/003465304323023750)
13. Fairlie RW, Hoffmann F, Oreopoulos P. 2011 *A community college instructor like me: race and ethnicity interactions in the classroom*. Cambridge, MA: National Bureau of Economic Research.
14. Marx DM, Roman JS. 2002 Female role models: protecting women's math test performance. *Personal Soc. Psychol Bull.* **28**, 1183–1193. (doi:10.1177/01461672022812004)
15. McIntyre RB, Lord CG, Gresky DM, Ten Eyck LL, Frye GDJ, Bond Jr CF. 2005 A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. *Curr. Res. Soc. Psychol.* **10**, 116–136.
16. Schinske J, Cardenas M, Kaliangara J. 2015 Uncovering scientist stereotypes and their relationships with student race and student success in a diverse, community college setting. *CBE—Life Sci. Educ.* **14**, ar35.
17. Steinke J, Lapinski M, Long M, Van Der Maas C, Ryan L, Applegate B. 2009 Seeing oneself as a scientist: media influences and adolescent girls' science career-possible selves. *J. Women Minor Sci. Eng.* **15**, 279–301. (doi:10.1615/jwomenminorscieneng.v15.i4.10)
18. Schinske JN, Perkins H, Snyder A, Wyer M. 2016 Scientist spotlight homework assignments shift students' stereotypes of scientists and enhance science identity in a diverse introductory science class. *CBE—Life Sci. Educ.* **15**, ar47.
19. Damschen EI, Rosenfeld KM, Wyer M, Murphy-Medley D, Wentworth TR, Haddad NM. 2005 Visibility matters: increasing knowledge of women's contributions to ecology. *Front. Ecol. Environ.* **3**, 212–219. (doi:10.1890/1540-9295(2005)003[0212:VMIKOW]2.0.CO;2)
20. Phillips J, Hausbeck K. 2000 Just beneath the surface: rereading geology, rescripting the knowledge-power nexus. *Women's Stud. Q.* **28**, 181–202.
21. Hooks B. 1984 *Feminist theory: from margin to center*. Cambridge, MA: South End Press.
22. Ong M, Wright C, Espinosa L, Orfield G. 2011 Inside the double bind: a synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics. *Harv. Educ. Rev.* **81**, 172–209. (doi:10.17763/haer.81.2.t022245n7×4752v2)
23. Chen X. 2013 *STEM attrition: college students' paths into and out of STEM fields*. Washington, DC: National Center for Education Statistics.
24. Mervis J. 2010 Better intro courses seen as key to reducing attrition of STEM majors. *Science* **330**, 306. (doi:10.1126/science.330.6002.306)
25. Seymour E, Hewitt NM. 1997 *Talking about leaving: why undergraduates leave the sciences*. Boulder, CO, USA: Westview Press.
26. Olson S, Riordan DG. 2012 *Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: Executive Office of the President.
27. Ballen CJ, Greene HW. 2017 Walking and talking the tree of life: why and how to teach about biodiversity. *PLoS Biol.* **15**, e2001630. (doi:10.1371/journal.pbio.2001630)
28. National Institutes of Health. 2015 Racial and ethnic categories and definitions for NIH diversity programs and for other reporting purposes. See https://grants.nih.gov/grants/guide/notice-files/not-od-15-089.html.
29. NSF. 2014 *Science and engineering indicators 2014*. Alexandria, VA: National Science Foundation.
30. Chou M. 2015 The Rise of the Young Professor. *Phys. World.* **13**, 57.
31. Bates D, Mächler M, Bolker B, Walker S. 2014 Fitting linear mixed-effects models using lme4. *arXiv* 14065823.
32. R Core Team. 2019 *R: A language and environment for statistical computing*. R version 3.6. 0. Vienna, Austria: R Foundation for Statistical Computing. See https://www.r-project.org/.
33. Fox J, Weisberg S. 2011 *Multivariate linear models in R: an R companion to applied regression*. Thousand Oaks, CA: Sage.
34. Growth AP, Distribution G, Data HO. 2011 Newsroom. See https://www.census.gov/newsroom.html.
35. Snyder TD, de Brey C, Dillow SA. 2019 *Digest of education statistics 2017*. Washington, DC: National Center for Education Statistics.
36. Landivar LC. 2013 Disparities in STEM employment by sex, race, and Hispanic origin. *Educ. Rev.* **29**, 911–922.
37. Henning JA, Ballen CJ, Molina S, Cotner S. 2019 Hidden identities shape student perceptions of active learning environments. *Front. Education* **4**, 129. (doi:10.3389/feduc.2019.00129)
38. Gibson DE. 2004 Role models in career development: new directions for theory and research. *J. Vocat. Behav.* **65**, 134–156. (doi:10.1016/S0001-8791(03)00051-4)
39. Lockwood P. 2006 'Someone like me can be successful': do college students need same-gender role models? *Psychol. Women Q.* **30**, 36–46. (doi:10.1111/j.1471-6402.2006.00260.x)
40. Koberg CS, Boss RW, Goodman E. 1998 Factors and outcomes associated with mentoring among health-care professionals. *J. Vocat. Behav.* **53**, 58–72. (doi:10.1006/jvbe.1997.1607)
41. Hurtado S, Newman CB, Tran MC, Chang MJ. 2010 Improving the rate of success for underrepresented racial minorities in STEM fields: insights from a national project. *New Dir. Institutional Res.* **2010**, 5–15. (doi:10.1002/ir.357)
42. Ceci SJ, Williams WM. 2011 Understanding current causes of women's underrepresentation in science. *Proc. Natl Acad. Sci.* **108**, 3157–3162. (doi:10.1073/pnas.1014871108)
43. Cotner S, Ballen C, Brooks DC, Moore R. 2011 Instructor gender and student confidence in the sciences: a need for more role models. *J. Coll Sci. Teach.* **40**, 96–101.
44. Williams J, Phillips KW, Hall EV. 2014 *Double jeopardy? Gender bias against women of color in science*. San Francisco, CA: Hastings College of the Law.
45. Greenfieldboyce N. 2019 Academic science rethinks all-too-white 'dude walls' of honor. See https://www.npr.org/sections/health-shots/2019/08/25/749886989/academic-science-rethinks-all-too-white-dude-walls-of-honor.