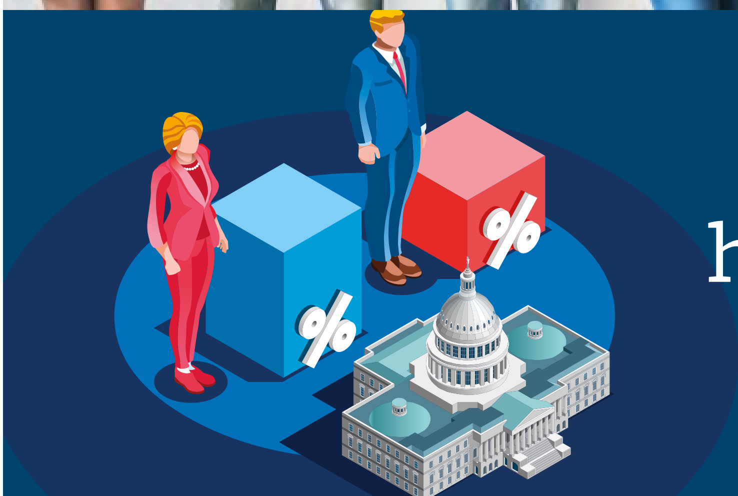
TNIT News
TOULOUSE NETWORK FOR INFORMATION TECHNOLOGY

Geek tragedy: Why do IT firms lack diversity?

Glenn Ellison
(MIT, NBER and TNIT)



Did fake news help Trump win?

Matthew Gentzkow
(Stanford)

➔ Issue 18 ➔ February 2018

Toulouse School of Economics
IDEI Institut d'Économie Industrielle
Microsoft

Strength in diversity

Jacques Crémer & Priyanka Talim

Diversity of backgrounds, ideas and voices have fueled the political and economic success of US democracy.

Technological innovation has been a key part of that story, but recently tech firms have come under attack for failing to strike the right balance.

In this issue of TNIT News, **MIT economist Glenn Ellison** discusses the technology sector's increasingly conspicuous lack of ethnic, gender and social diversity. IT firms have plenty of work to do, and much to gain by increasing the involvement of women and minorities. Glenn's research highlights an important component of the problem that is often overlooked: **the need to promote and diversify computer science education in schools and colleges.**

Also in these pages, **Stanford economist Matthew Gentzkow** looks at the role of social media and fake news in the 2016 US presidential election. Have social networks made the "marketplace of ideas" more diverse and competitive? Or is IT to blame for the segregation of voters into "echo chambers" or "filter bubbles", vulnerable to manipulation via algorithmic targeting or botnets, and the spread of misinformation?
Matthew's analysis represents a timely addition of hard evidence and academic rigor to the debate.

Geek tragedy: Why do IT firms lack diversity?

by Glenn Ellison
(MIT, NBER and TNIT)

The underrepresentation of women and minorities in the IT workforce is a longstanding concern. IT firms lose the potential benefits of a diverse workplace and the failure to attract talented women and minorities contributes to the overall shortage of highly trained workers.

The 21st century has seen strikingly little progress to diversify the IT workforce, mirroring a lack of progress in education. Substantial progress will likely require sustained effort on multiple fronts.

Only about one in four US workers in mathematical or computer science are female, and this fraction has held strikingly steady since the early 2000s *(see Figure 1)*. Recent increases in minority representation are encouraging, but much of the longer trend is at best just keeping up with the changing composition of the US labor force. For example, between 2000 and 2015 the proportion of Hispanics working in mathematical or computer science increased from 5.1% to 6.8%, but at the same time Hispanics in the full US workforce increased from 12.1% to 16.4%.

Computer Science and Math Professionals:
Fraction Female, Black, and Hispanic

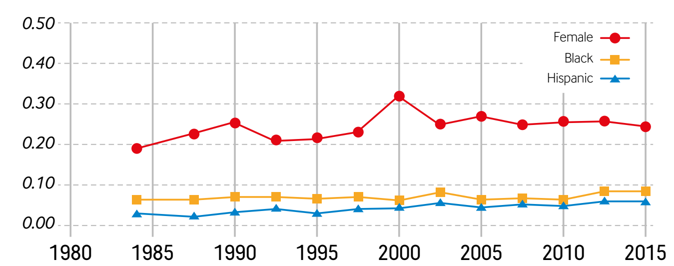


Figure 1

Looking back: diversity in college programs

The lack of success in diversifying IT workplaces mirrors another lack-of-success story: there has been little progress in diversifying computer science programs at US universities *(see Figure 2)*. Black representation rose in the 1990s but declined in the past decade. Hispanic representation has had a slow but steady increase.

Each group now comprises about 10% of recent bachelor's degrees in computer science. But both groups remain underrepresented and the increase in Hispanic representation just follows the trend in the overall college population. More strikingly, female representation was up to 37% by 1984, but the past 30 years have seen a long decline to about half of the peak level. This decline was noticeably steep in the late 1980s and mid 2000s.

There is a vast literature on women's underrepresentation in STEM fields (science, technology, engineering and mathematics). A wide variety of potential causes have been explored, including biological differences, childhood influences, treatment by teachers, reluctance to enter male-dominated fields, biases in evaluation and family demands. A wealth of evidence illustrates the relevance of many of these mechanisms.

The outcomes in college computer science, however, are not like those of other STEM fields. In the 1970s and early 1980s the trends were similar. Women were gaining ground in many male-dominated fields, reaching near parity in biology and mathematics by the mid-1980s. Computer science diverged from the path at this point. In engineering and the physical sciences, the proportion of female BAs continued to rise for another 20

Computer Science BAs:
Fraction Female, Black, and Hispanic

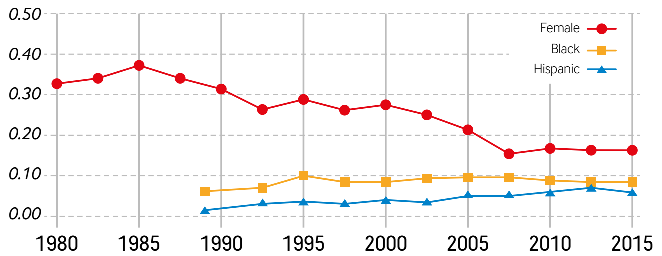


Figure 2

years after that of female computer science majors started to fall *(see Figure 3)*. In the past decade the trends have again become similar, with all three fields experiencing a leveling off.

Why has the experience of computer science been so different? Some have pointed to the early 1980s as a time when popular culture developed an image of the single-minded, poorly socialized computer geek that jarred with many women's self-image and aspirations. It is also perhaps noteworthy that students started to arrive on campus with experience of programming PCs in the mid 1980s; and cohorts graduating in the mid 2000s were the first to have completed high school after the dot-com boom. Young women have been noted to have lower self-assessments of their mathematical ability than comparably accomplished young men, and this can deter them from entering technical fields. Increased pre-college exposure to computers could have shifted the freshman experience from one where male and female students arrived on campus equally inexperienced in computers to one where freshman women were or felt behind.

Looking further back: diversity in advanced high-school work

Even apart from self-confidence effects, students are more likely to pursue and persist in majors for which they are well prepared in high school. It is presumably very relevant that fewer women and minorities enter college with strong computer science backgrounds. For example, in 2014 the population of students with passing grades (3+) on the AP Computer Science A exam was just 18.7% female, 2.6% black and 5.6% Hispanic.

The underrepresentation of minorities is not hard to understand: access to high-quality high-school computer science classes is limited. In 2010 less than one in ten US high schools had any student take the AP Computer Science exam, and computer science classes are disproportionately found in schools serving students of high socioeconomic status.

> **Women were gaining ground in many male-dominated fields, reaching near parity in biology and mathematics by the mid-1980s**

In a recent joint paper with Parag Pathak, we note that a number of US public school systems have recently abandoned race-based affirmative action admissions policies. Given that many large urban systems have traditionally offered advanced classes only in a few magnet schools, this may be an additional obstacle to increasing access for talented minority students.

The male-female gap in high-school computer science course-taking is not directly attributable to girls not having access. The gender gap in AP course-taking is much larger in some STEM subjects than others *(see Figure 4)*. Girls achieve the majority of passing scores on AP Biology and over 40% of them on AP Chemistry, but comprise a much smaller fraction of those taking AP Physics and Computer Science.

Why are outcomes so different? One potentially relevant observation is that the AP courses with smaller gender gaps are taken by many more students. A first biology course is usually required of all students and taken early in high school. Courses in computer science and mathematically-rigorous physics courses are usually not required and are taken later, if at all. High-school girls in general are doing very well: they get better grades than boys; take at least as many advanced courses, etc. Different expectations about course choices may make a very big difference in the gender gap.

A paper of mine with Ashley Swanson (Ellison and Swanson (2010)) finds that the gender gap in high-school mathematics competitions is larger than the gender gap in SAT scores when we look at comparably high performance levels. This indicates that high-ability girls are choosing not to participate. When we look at extremely high performance levels, the gender gap is

Fraction of BAs Female: Computer Science vs. Physical Sciences and Engineering

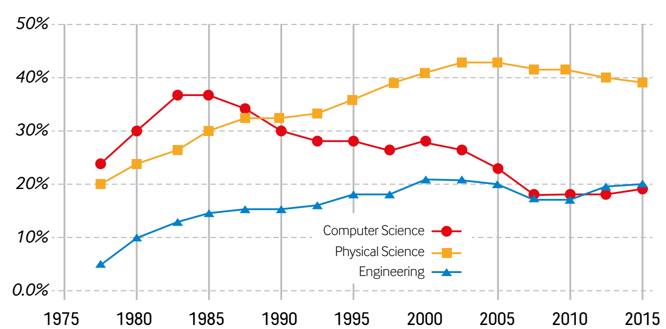


Figure 3

smaller at the highest-achieving high schools. One explanation is that more girls at the high-achieving schools may be pursuing advanced math because they are part of a like-minded community. This suggests that the gender gap in advanced computer science may narrow if the subject becomes more common.

Recent efforts in high-school computer science

Microsoft, Google and other tech firms have supported a number of efforts to bolster high-school computer science education. This includes both out-of-school programs like Girls Who Code and Made with Code, and in-school programs like Microsoft's TEALS, which connects high schools with volunteers and has expanded from four schools in 2010-2011 to 162 schools in 2015-2016.

Fraction Female vs Popularity of AP Science Tests

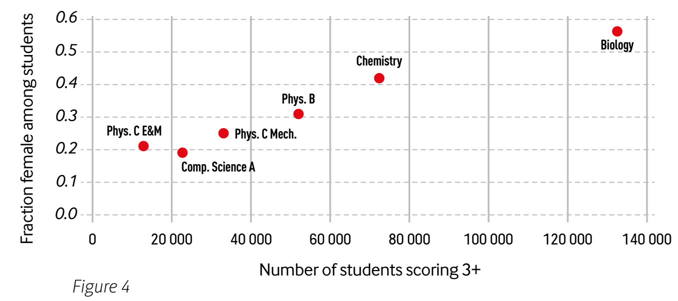


Figure 4

Overall AP participation has recently grown (46% more tests were taken in 2016 than in 2010), although by 2010 AP Computer Science participation seemed overdue for an increase – it had not grown at all between 2000 and 2010 whereas overall AP test-taking increased by 153%. However, there are substantial challenges. Many schools offer no computer science classes at all, and many schools lack teachers with sufficient expertise. In schools that newly offer AP Computer Science, the results are often quite poor.

Finally, we come to some good news. The growth of AP Computer Science since 2010 has been a great success story. The number of students taking the exam has nearly tripled in just six years. The number of students with passing marks has grown nearly as fast.

The gender gap in AP Computer Science has narrowed: the percentage female among students with passing marks increased from 17.5% in 2010 to 22.1% in 2016. Hispanic representation also increased from 4.9% in 2010 to 7.5% in 2016. The percentage of black students increased slightly, but then fell back to the 2010 level. Again, this may reflect the greater challenges of improving minority outcomes: many schools serving minority communities lack the physical and/or human resources needed to offer computer science courses; and we do not yet know how restrictions on affirmative action have affected access to magnet programs.

While the percentage increase numbers are not so striking, looking at the number of passing students from each group is more encouraging: the number of black, Hispanic and female students with passing marks on the AP Computer Science exam have each more than tripled in just six years. The 7,674 girls who passed an AP Computer Science exam in 2016 is a very small number for a country the size of the US and pales in comparison with roughly 80,000 girls passing AP Biology, but it is much more encouraging than just 2,201. Minority counts remain very low in absolute terms.

SUMMING UP

The loss of undergraduate women since 1985 sets computer science apart from other STEM fields, and there have been only small improvements in minority representation. Diversifying the IT workforce will require a variety of efforts at the employment, retention, and advancement stages, but IT firms must look beyond their own practices. A large effort to improve undergraduate outcomes will also be required.

One factor contributing to low female and minority representation at college level is the disparity in advanced high-school work. The challenges of increasing female and minority representation are different. Computer science courses (let alone high-quality ones) are simply unavailable in the high schools that many minority students attend. Increasing female participation is probably a case of making computer science a normal part of any student's program – high-school girls are doing very well in in many other areas where their success is expected.

The recent surge in AP Computer Science is an indication that some progress is being made. The gains in diversity in advanced high-school coursework are not yet very large. And it will be a few years before we will see if they are followed by improved diversity at colleges. But it is encouraging that concerted efforts have helped produce a large surge in participation accompanied by some progress in diversity.

Did fake news help Trump win?

by Matthew Gentzkow
(Stanford)

American democracy has been repeatedly buffeted by changes in media technology. From the introduction of cheap "penny papers" in the 19th century, to the rise of radio and television, to the early days of the internet, new media have both produced profound effects, and provoked outsized anxieties.

Following the 2016 election, the focus of concern has shifted to social media. Social media has rapidly become one of the most important channels by which people consume political news and information. This may have important upsides for democracy. Social networks have engaged a large swath of voters who do not consume traditional media, allowed more direct communication between politicians and voters, and, by most accounts, made the "marketplace of ideas" more diverse and competitive.

However, as with all previous media technologies, the potential downsides loom far larger, at least in the public imagination. These include segregation of voters into "echo chambers" or "filter bubbles", vulnerability to manipulation by bad actors via algorithmic targeting or botnets, and above all the proliferation of misinformation or, as the ubiquitous cliché would have it, "fake news".

Research teams in academia, non-profits, and private firms have begun to quantify the scale of these risks, and to evaluate potential solutions. Though the evidence remains fragmentary, this collective effort provides some much-needed discipline to the broader conversation.

In a recent paper with NYU economist Hunt Allcott[1], we offer one contribution in this vein, combining audience data, fact-checking archives, and a new, specially commissioned, 1,200-person post-election survey to gauge the reach of fake news in the run-up to the 2016 election and the economic forces at play in its distribution.

We define fake news to be news articles that are intentionally and verifiably false, with the potential to mislead readers. This definition includes intentionally fabricated news stories, such as a widely shared article from the now-defunct website denverguardian.com with the headline "FBI agent suspected in Hillary email leads found dead in apparent murder-suicide". It also includes many articles that originate on satirical websites but could be misunderstood as factual when viewed in isolation on Facebook or Twitter. For example, in July 2016, the now defunct wtoe5news.com reported that Pope Francis had endorsed Donald Trump's presidential candidacy. According to the site's "about" page, "Most articles on wtoe5news.com are satire or pure fantasy." But this disclaimer was not included in the article itself, and the story was shared more than one million times on Facebook.

Our definition rules out several close cousins of fake news: (1) unintentional reporting mistakes, (2) rumors that do not originate from a particular news article, (3) conspiracy theories spread by people who believe them to be true, (4) satire unlikely to be misconstrued, (5) false statements by politicians, and (6) reporting that is slanted or misleading but not outright false.

As a first step in our analysis, we sketched a stylized theoretical model of the market for news on social media. The model clarifies the key forces that lead fake news to arise and proliferate: low costs of production and distribution, difficulty for consumers in separating false from true stories, tastes on the part of consumers for partisan confirmation, and large economic returns from advertising. The model also highlights the fact that the potential social costs include not just biased beliefs about the fake news stories at issue, but also a reduction in trust in media outlets more generally.

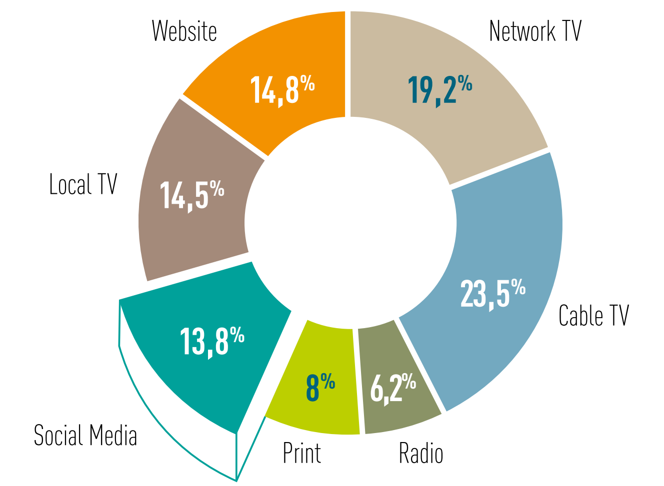
> **Education, age, and total media consumption are strongly associated with more accurate beliefs about whether headlines are true or false**

Informed by this model, we then studied the role of fake news in the 2016 American presidential election.

First, we assessed the importance of social media relative to other sources of political news and information. Figure 1 shows the media which respondents to our survey described as their "most important" sources of news about the 2016 election. Television remains by far the most influential source, chosen as most important by a majority of voters. Only 14 percent of voters cite social media as their most important source. We conclude that social media was important, but far from dominant, and probably less important than some might infer from its prominence in the public discussion post-election.

We also assembled a database of 156 election-related news stories that were categorized as false by leading fact-checking websites in the three months before the election. This enabled us to confirm that fake news was both widely shared and heavily tilted in favor of Donald Trump. Our database contains 115 pro-Trump fake stories that were shared on Facebook a total of 30 million times, and 41 pro-Clinton fake stories shared a total of 7.6 million times.

Several benchmarks allowed us to measure the rate at which voters were exposed to fake news. The upper end of previously reported statistics for the ratio of page visits to shares of stories on Facebook would suggest that each of the 38 million shares of fake news in our database lead each to on average 20 clicks on the websites posting the stories. This translates into 760 million instances of a user clicking through and reading a fake news story, or about three stories read per American adult. A list of fake news websites, on which just over half of articles appear to be false, received 159 million visits during the month of the election, or 0.64 per US adult. In our post-election survey, about 15 percent of respondents recalled seeing each of 14 major pre-election fake news headlines, but about 14 percent also recalled seeing a set of placebo fake news headlines - untrue headlines that we invented and that never actually circulated. Using the difference between "real" fake news headlines and our placebos as a measure of true recall and projecting this to the universe of fake news articles in our database, we estimate that the average adult saw and remembered 1.14 stories. Taken together, these estimates suggest that the average US adult might have seen perhaps one or several (but not ten or a hundred) fake news stories in the months before the election.



Finally, we studied the ability of our survey respondents to distinguish false and true headlines. Education, age, and total media consumption are strongly associated with more accurate beliefs about whether headlines are true or false. Democrats and Republicans are both about 15 percent more likely to believe ideologically aligned headlines, and this ideologically tilted inference is stronger for those with more ideologically segregated social media networks.

This analysis is not sufficient to allow us to say anything conclusive about the potential impact of fake news on the 2016 election outcome, because we do not have an estimate of the way fake news exposure affected votes. However, back-of-the-envelope calculations suggest that seeing a single fake news story would have needed to be many times more persuasive than seeing a single TV commercial in order for fake news to have changed the election outcome.

REFERENCES

Note: I included reference numbers in square brackets in the text rather than standard author-date references. My thought was that when the piece is "published" these could all be replaced with hyperlinks and the actual references omitted.

[1] Allcott, Hunt and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. Journal of Economic Perspectives. 31(2): 211-236

TNIT News

TOULOUSE NETWORK FOR INFORMATION TECHNOLOGY

- *Scientific Director:* Jacques Crémer
- *Production Manager:* Priyanka Talim
- *Editorial contributions:* James Nash
- *Graphics:* Olivier Colombe
- *Illustrations:* I-Stock

TNIT
Manufacture des Tabacs
21 allée de Brienne
31015 Toulouse Cedex 6 France
+33(0)5 61 12 85 89

www.idei.fr

tnit@tse-fr.eu

The Toulouse Network for Information Technology (TNIT) is a research network funded by Microsoft and managed by the Institut d'Economie Industrielle (IDEI). It aims at stimulating world-class research in the economics of information technology, intellectual property, software security, liability, and related topics.

➔ Issue 18 ➔ March 2018

Toulouse School of Economics

IDEI — Institut d'Economie Industrielle

Microsoft