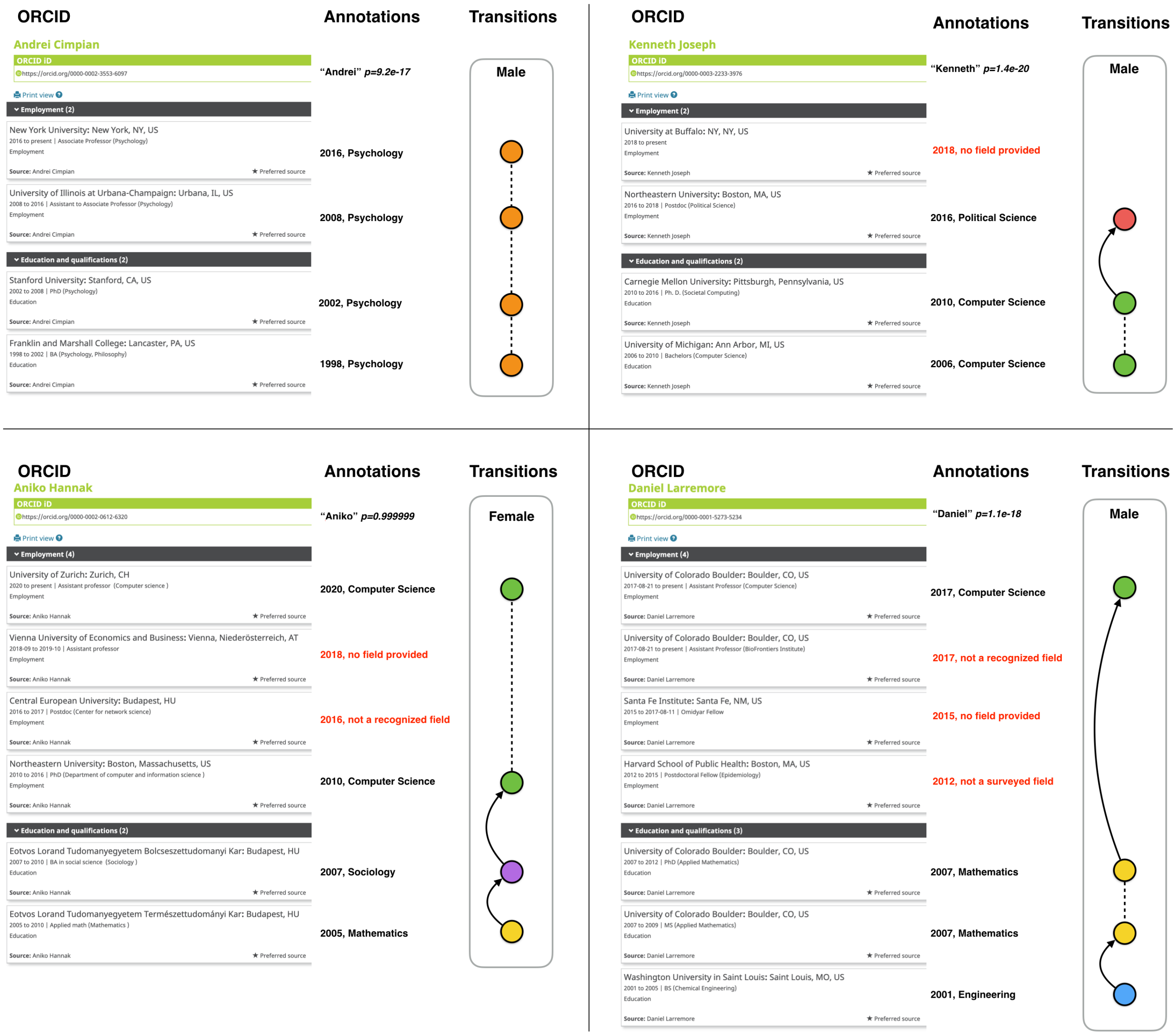
# Supplementary Online Materials

## Field-specific Ability Beliefs as an Explanation for Gender Differences in Academics’ Career Trajectories: Evidence from Public Profiles on ORCID.org



Andrei Cimpian

ORCID iD

https://orcid.org/0000-0002-3553-6097

Print view

Employment (2)

New York University: New York, NY, US
2016 to present | Associate Professor (Psychology)
Employment
**Source:** Andrei Cimpian ★ Preferred source

University of Illinois at Urbana-Champaign: Urbana, IL, US
2008 to 2016 | Assistant to Associate Professor (Psychology)
Employment
**Source:** Andrei Cimpian ★ Preferred source

Education and qualifications (2)

Stanford University: Stanford, CA, US
2002 to 2008 | PhD (Psychology)
Education
**Source:** Andrei Cimpian ★ Preferred source

Franklin and Marshall College: Lancaster, PA, US
1998 to 2002 | BA (Psychology, Philosophy)
Education
**Source:** Andrei Cimpian ★ Preferred source

Kenneth Joseph

ORCID iD

https://orcid.org/0000-0003-2233-3976

Print view

Employment (2)

University at Buffalo: NY, NY, US
2018 to present
Employment
**Source:** Kenneth Joseph ★ Preferred source

Northeastern University: Boston, MA, US
2016 to 2018 | Postdoc (Political Science)
Employment
**Source:** Kenneth Joseph ★ Preferred source

Education and qualifications (2)

Carnegie Mellon University: Pittsburgh, Pennsylvania, US
2010 to 2016 | Ph. D. (Societal Computing)
Education
**Source:** Kenneth Joseph ★ Preferred source

University of Michigan: Ann Arbor, MI, US
2006 to 2010 | Bachelors (Computer Science)
Education
**Source:** Kenneth Joseph ★ Preferred source

Aniko Hannak

ORCID iD

https://orcid.org/0000-0002-0612-6320

Print view

Employment (4)

University of Zurich: Zurich, CH
2020 to present | Assistant professor (Computer science )
Employment
**Source:** Aniko Hannak ★ Preferred source

Vienna University of Economics and Business: Vienna, Niederösterreich, AT
2018-09 to 2019-10 | Assistant professor
Employment
**Source:** Aniko Hannak ★ Preferred source

Central European University: Budapest, HU
2016 to 2017 | Postdoc (Center for network science)
Employment
**Source:** Aniko Hannak ★ Preferred source

Northeastern University: Boston, Massachusetts, US
2010 to 2016 | PhD (Department of computer and information science )
Employment
**Source:** Aniko Hannak ★ Preferred source

Education and qualifications (2)

Eotvos Lorand Tudomanyegyetem Bolcseszettudomanyi Kar: Budapest, HU
2007 to 2010 | BA in social science (Sociology )
Education
**Source:** Aniko Hannak ★ Preferred source

Eotvos Lorand Tudomanyegyetem Természettudományi Kar: Budapest, HU
2005 to 2010 | Applied math (Mathematics )
Education
**Source:** Aniko Hannak ★ Preferred source

Daniel Larremore

ORCID iD

https://orcid.org/0000-0001-5273-5234

Print view

Employment (4)

University of Colorado Boulder: Boulder, CO, US
2017-08-21 to present | Assistant Professor (Computer Science)
Employment
**Source:** Daniel Larremore ★ Preferred source

University of Colorado Boulder: Boulder, CO, US
2017-08-21 to present | Assistant Professor (BioFrontiers Institute)
Employment
**Source:** Daniel Larremore ★ Preferred source

Santa Fe Institute: Santa Fe, NM, US
2015 to 2017-08-11 | Omidyar Fellow
Employment
**Source:** Daniel Larremore ★ Preferred source

Harvard School of Public Health: Boston, MA, US
2012 to 2015 | Postdoctoral Fellow (Epidemiology)
Employment
**Source:** Daniel Larremore ★ Preferred source

Education and qualifications (3)

University of Colorado Boulder: Boulder, CO, US
2007 to 2012 | PhD (Applied Mathematics)
Education
**Source:** Daniel Larremore ★ Preferred source

University of Colorado Boulder: Boulder, CO, US
2007 to 2009 | MS (Applied Mathematics)
Education
**Source:** Daniel Larremore ★ Preferred source

Washington University in Saint Louis: Saint Louis, MO, US
2001 to 2005 | BS (Chemical Engineering)
Education
**Source:** Daniel Larremore ★ Preferred source

**Figure S1:** Public ORCID profiles for the authors of the study, with annotations illustrating choices made during data processing.

**Table S1:** Breakdown of ORCID users in our dataset by field and gender (proportions in parentheses).

| | **Everyone, regardless of whether they ever switched fields** | | | **Those who switched out of these fields at some point** | | |
|---|---|---|---|---|---|---|
| Field | Total | Men | Women | Total | Men | Women |
| Anthropology | 10335 | 4920 (0.48) | 5415 (0.52) | 1734 | 832 (0.48) | 902 (0.52) |
| Archaeology | 4432 | 2462 (0.56) | 1970 (0.44) | 755 | 389 (0.52) | 366 (0.48) |
| Art History | 3165 | 1223 (0.39) | 1942 (0.61) | 531 | 200 (0.38) | 331 (0.62) |
| Astronomy* | 62 | 47 (0.76) | 15 (0.24) | 13 | 9 (0.69) | 4 (0.31) |
| Biochemistry* | 22020 | 13402 (0.61) | 8618 (0.39) | 2892 | 1887 (0.65) | 1005 (0.35) |
| Chemistry* | 84639 | 57127 (0.67) | 27512 (0.33) | 10198 | 7219 (0.71) | 2979 (0.29) |
| Classics | 2371 | 1212 (0.51) | 1159 (0.49) | 538 | 265 (0.49) | 273 (0.51) |
| Communications | 20570 | 10162 (0.49) | 10408 (0.51) | 2918 | 1525 (0.52) | 1393 (0.48) |
| Comparative Literature | 996 | 457 (0.46) | 539 (0.54) | 252 | 114 (0.45) | 138 (0.55) |
| Computer Science* | 50142 | 39225 (0.78) | 10917 (0.22) | 6158 | 4939 (0.80) | 1219 (0.20) |
| Earth Sciences* | 14185 | 9557 (0.67) | 4628 (0.33) | 1504 | 1011 (0.67) | 493 (0.33) |
| Economics | 54297 | 34725 (0.64) | 19572 (0.36) | 4154 | 2665 (0.64) | 1489 (0.36) |
| Education | 78997 | 35391 (0.45) | 43606 (0.55) | 8570 | 4607 (0.54) | 3963 (0.46) |
| Engineering* | 227769 | 180918 (0.79) | 46851 (0.21) | 17694 | 14170 (0.80) | 3524 (0.20) |
| English Literature | 19345 | 8769 (0.45) | 10576 (0.55) | 3730 | 1682 (0.45) | 2048 (0.55) |
| Evolutionary Biology* | 3024 | 1707 (0.56) | 1317 (0.44) | 178 | 89 (0.50) | 89 (0.50) |
| History | 25840 | 15144 (0.59) | 10696 (0.41) | 4039 | 2322 (0.57) | 1717 (0.43) |
| Linguistics | 9857 | 4332 (0.44) | 5525 (0.56) | 1940 | 918 (0.47) | 1022 (0.53) |
| Mathematics* | 42416 | 31305 (0.74) | 11111 (0.26) | 8770 | 6513 (0.74) | 2257 (0.26) |
| Middle Eastern Studies | 568 | 373 (0.66) | 195 (0.34) | 135 | 85 (0.63) | 50 (0.37) |
| Molecular Biology* | 9464 | 5429 (0.57) | 4035 (0.43) | 1071 | 678 (0.63) | 393 (0.37) |
| Music Theory & Composition | 1317 | 740 (0.56) | 577 (0.44) | 196 | 110 (0.56) | 86 (0.44) |
| Neuroscience* | 12552 | 6947 (0.55) | 5605 (0.45) | 1154 | 629 (0.55) | 525 (0.45) |
| Philosophy | 18098 | 11866 (0.66) | 6232 (0.34) | 3630 | 2281 (0.63) | 1349 (0.37) |
| Physics* | 81319 | 64817 (0.80) | 16502 (0.20) | 12552 | 10209 (0.81) | 2343 (0.19) |
| Political Science | 18517 | 11555 (0.62) | 6962 (0.38) | 2457 | 1482 (0.60) | 975 (0.40) |
| Psychology | 57890 | 23920 (0.41) | 33970 (0.59) | 6460 | 2940 (0.46) | 3520 (0.54) |
| Sociology | 17374 | 8930 (0.51) | 8444 (0.49) | 2605 | 1353 (0.52) | 1252 (0.48) |
| Spanish Literature | 2365 | 1035 (0.44) | 1330 (0.56) | 544 | 237 (0.44) | 307 (0.56) |
| Statistics* | 11288 | 7433 (0.66) | 3855 (0.34) | 1679 | 1177 (0.70) | 502 (0.30) |

*Note.* The 12 fields marked with an asterisk were classified as STEM; the other 18 were non-STEM [1]. The two sets of statistics above correspond to Step 2 and 3 in ORCID data processing sequence described in Section 1. However, the numbers in the “Total” columns will add up to more than the numbers provided in Section 1 because users can be affiliated with more than one field. The sets of numbers on the left (the ”Everyone” columns) include any ORCID users who listed *at least one affiliation* in their profile. This is a larger sample than those in the analyses reported in the main text, which consisted of
(1) individuals with at least two affiliations who switched fields at least once (the recruitment analyses) or
(2) individuals with at least two affiliations regardless of whether or not they switched fields (the retention analyses).

# 1. ORCID Data Processing Information

We provide a narrative summary of our data processing steps to accompany our publicly available scripts at https://github.com/kennyjoseph/ORCID_trajectories.

ORCID users have the option to selectively make their information public. This opt-in public information is released annually in an aggregated ORCID public data file [2] and also made available on demand to ORCID member institutions. The most recent version of ORCID public data were accessed on February 18, 2021.

Our goal is to identify the subset of researchers whom we can confidently place in particular fields of study, over time or career stage, and whose first and last names provide us with a high-confidence association with a gender. However, ORCID does not provide information about its users' fields of study or gender. As a consequence, not all ORCID profiles could be analyzed, so this section describes our data cleaning and inclusion processes in detail. We begin with a dataset of 9,600,248 public profiles of researchers on ORCID. Only 3,006,777 researchers have one or more listed affiliations, and among them we observe a total of 8,146,175 affiliations.

**Step 1.** In the first step, we perform basic filtering to remove incomplete or out-of-scope affiliations. Specifically, we filter out three types of affiliations. First, we remove any affiliations where the researcher's name was not provided, as we use names to estimate an associated gender. Second, we filter out affiliations where the department name was not provided, as we use this information to identify the academic field associated with the affiliation. Finally, we remove any affiliations where neither a career stage (e.g., "postdoc") nor a date was provided, as we use this information to determine the ordering of affiliations. (The order of affiliations will later be used to identify field entries and exits.) After these three filters have been applied, we retain 5,722,977 affiliations among 2,381,829 researchers.

**Step 2.** In the second step, we fill in variables of interest using three algorithms: one to determine the role/job position affiliated with each affiliation (e.g., PhD candidate, professor; see Section 1-A), one to identify the academic field associated with each affiliation (see Section 1-B), and one to determine the gender that is culturally associated with researchers' names (see Section 1-D). Details on the algorithms themselves can be found in the sections that follow, but here we summarize their outputs. We remove affiliations that (i) match a field that is not among the 30 surveyed fields or (ii) match multiple fields, retaining 1,768,238 affiliations (965,603 researchers). Among these, we are able to find a high-confidence inferred gender via a cultural consensus algorithm for 1,480,407 affiliations (809,988 researchers; see Table S1 for a breakdown by field).

**Step 3.** In the third step, we take the 1,480,407 affiliations among 809,988 researchers and identify pairs of affiliations that indicate that a researcher exited one of the 30 surveyed fields and entered another (a "field switch" or "field transition"; see Section 1-E). To do so, we order each researcher's affiliations using the roles identified in Step 2 and/or the start date of the affiliation (for an illustration, see Figure S1). Whenever the fields associated with consecutive affiliations are different, we record a transition as having occurred from one affiliation's field to the next affiliation's field. If an individual changes fields multiple times, each transition is recorded as a separate transition, but the transitive transition (from the first field to the third field) is not recorded. In total, we are able to identify 112,132 transitions among 86,879 researchers (see Table S1), averaging 1.3 transitions per person among the researchers with observable transitions.

**Table S2:** Breakdown of career stages in the dataset.

| Role | Number of Affiliations | % of Data |
|---|---|---|
| Bachelor's Degree | 217667 | 12.3% |
| Master's Degree | 250772 | 14.2% |
| PhD | 402501 | 22.8% |
| Postdoctoral Researcher | 65324 | 3.7% |
| Professor/Department Head | 375565 | 21.2% |
| Other/None | 456409 | 25.8% |

## 1-A. Determining Career Stages

Each ORCID affiliation has an associated **role** field, which we use to identify the career stage associated with that affiliation. Due to the fact that the ORCID userbase spans many languages and academic traditions, we used a set of regular expressions to coarse-grain each affiliation into one of the following academic career stages: bachelor's degree, master's degree, PhD, postdoctoral researcher, and professor/department head. In the event that the text in an affiliation matches multiple stages, we select the highest ranking role. In the event that there is no match, we give that affiliation a blank role, since affiliations that have no role but nevertheless have a date that can be placed in sequence with other affiliations are still useful in our analysis.

The regular expressions were accumulated recursively: After every iteration of matching, we manually identified the most commonly missed expressions among unmatched roles, and then added a corresponding regular expression. We stopped once the inclusion of additional regular expressions did not substantially improve our data coverage. The complete set of regular expressions has been made publicly available. A breakdown of the career stages identified in our dataset via this algorithm can be found in Table S2.

## 1-B. Determining Academic Fields

Each ORCID affiliation has an associated **department name** field (hereafter, department). Only those affiliations

**Table S3:** Fields and the corresponding terms used for matching ORCID affiliation strings.

| Fields | Accepted expressions |
|---|---|
| Anthropology | anthropology |
| Archaeology | archaeology |
| Art History | art history; history of art |
| Astronomy | astronomy |
| Biochemistry | biochemistry |
| Chemistry | chemistry |
| Classics | classics; classical literature; classical humanities |
| Communications | communications; communication sciences; communication studies; communication |
| Comparative Literature | comparative literature |
| Computer Science | computer science; algorithms; computing; informatics |
| Earth Sciences | earth sciences; earth science; physical geography; oceanography; atmospheric sciences; volcano |
| Economics | economics; economic; econometrics; finance; economy |
| Education | education; pedagogy |
| Engineering | engineering; ingegneria; e.e.; e.c.e.; ingenierã a; cybernetics; telecommunication; telecommunications; telecommunication studies; electrical engineering; chemical engineering; electrical and computer engineering; biochemical engineering; biological engineering; neuroengineering; musical engineering; statistical engineering; physical engineering |
| English Literature | english literature; english |
| Evolutionary Biology | evolutionary biology |
| History | history |
| Linguistics | linguistics; linguistic |
| Mathematics | mathematics; math; geometry; algebra; number |
| Middle Eastern Studies | middle eastern studies; middle east |
| Molecular Biology | molecular biology |
| Music Theory & Comp. | music theory; musical composition; musicology; composition |
| Neuroscience | neuroscience |
| Philosophy | philosophy |
| Physics | physics |
| Political Science | political science; political sciences; politics; science politique; politology |
| Psychology | psychology; psychological; psicologia; psicología |
| Sociology | sociology; sociological; sociologie |
| Spanish Literature | spanish literature; spanish |
| Statistics | statistics; statistical sciences |

that can be confidently linked to one and only one of the 30 surveyed academic fields [1] can be used in our analysis, so we now describe the procedure used to match user-provided departments with surveyed academic fields. There are three steps in our approach: translation, matching, and multi-field affiliation removal.

**Step 1: Translation.** In the translation step, we use a list of common academia-related English words to determine whether or not an affiliation's department is in English. We then translated the 232,960 non-English unique department names into English using Google Translate.

**Step 2: Matching.** In the matching step, we first construct, for each of the 30 surveyed fields, a list of expressions and subfields that are associated with that field. For example, *sociology* or *sociological* could both map to the field of sociology. Table S3 provides a complete list of fields and their corresponding expressions. Note that we retained certain popular non-English terms, as there were some instances in which certain terms were not translated (they were considered to be misspellings by the translation algorithm).

We also assembled a so-called *denylist* of scientific fields that are prominent in the ORCID data but that are not in our survey data. Constructing this list helped identify affiliations in fields that were clearly defined but outside the scope of our study; researchers who will use this dataset in the future might find these fields useful.

To construct the terms in Table S3 and our denylist, we used a recursive approach, using existing resources from Wikipedia and the U.S. National Science Foundation to determine initial lists of terms, and then repeatedly inspecting department names that appeared multiple times in our dataset to ensure coverage of our lists.

To validate the output of the matching step, we hand-checked the fields assigned to departments from a stratified random sample, consisting of 25% matched and in-sample affiliations, 25% unmatched affiliations, and 50% matched but out of sample (denylist) affiliations. Each affiliation was assigned to two of the four authors of the study for annotation. We used disagreements between annotators and the matching step to improve the expressions and denylist. In total, 660 affiliations were checked by hand by at least two authors, and all disagreements were discussed by all authors.

With the final set of terms associated with each field (see Table S3) and the denylist, we use a simple rule-based algorithm to match affiliations to fields. The algorithm works as follows.

First, it splits each affiliation string on common separators (e.g., commas, "and") into candidate match objects. For instance, consider the fake and implausible department (with intentional misspelling) *Advanced Chemical Engineering and Histroy/History of Art*. Based on the matching terms, this string contains three candidate match objects: (i) "Advanced Chemical Engineering", (ii) "Histroy", and (iii) "History of Art".

Then, for each candidate match object, we check whether it is an exact match to any term associated with a surveyed field or a field on the denylist. If so, we have identified the field associated with the candidate match object, and move to the next candidate match object. For example, "History of Art" matches a term in Table S3 associated with the field Art History.

If a candidate match object matches no known terms, we check for fuzzy string matches with an edit distance of 3 or less. If there are any such fuzzy matches, we select the one with the smallest edit distance and break ties by choosing the longest of the matched strings. For example, the candidate match object "Histroy" has an edit distance of 1 to the match term "History," which is linked to the academic field History.

Finally, we check for exact matches and/or fuzzy matches in subsets of the candidate match strings. For example, there is an exact match to "Chemical Engineering", a term in Table S3, within the candidate match object "Advanced Chemical Engineering". In such a case, we identify this field as a match, remove the relevant substring, and then continue recursively (i.e., try to match the remaining substring "Advanced"). Fuzzy matching was conducted in part using the *fuzzywuzzy* open source library.[1]

This field matching method identifies zero, one, or more academic fields associated with each affiliation. In total, 59.1% of the in-sample affiliations matched *exactly* to one or more of the terms in Table S3, with another 33.7% within an edit distance of 2 to one or more of those terms. Thus, 92.8% of the affiliations we identified as representing one of the 30 fields in our survey data were linked to that field either because they were an exact match to a term in Table S3 or were a slight misspelling of one of those terms.

**Step 3: Multi-field affiliation removal.** The matching procedure has the potential to associate multiple fields with a single affiliation. In such cases, the affiliation was conservatively removed from further consideration in order to avoid ambiguities.

### 1-C. Determining Region

Each ORCID affiliation has an associated *ISO 3166-1 alpha-2* country code, allowing us to test the extent to which field transitions in different parts of the world show the same patterns as those in the global dataset. Table S4 lists which countries are assigned to which regions, and the number of transitions to organizations in countries in that region.

### 1-D. Associating Names with Gender

ORCID neither collects nor infers gender information. Thus, we inferred the extent to which each user's first and last names are culturally associated with different gender labels. This inferential process was guided by the theoretical framework of cultural consensus models [3], which do not purport to identify the "true" gender label of an individual but instead measure the consensus across multiple viewpoints. In other words, this method does not ask "What is Jane Doe's gender?" but rather "What is the likelihood that someone with the name 'Jane Doe' is thought to be a woman?" In this way, the inference algorithm attempts to estimate how an individual is likely to be perceived based on their name.

Our consensus-based gender inference algorithm computes the Bayesian posterior probability that a person's name is culturally understood to be the name of a woman or a man based on data from 30 different sources. These sources range from the U.S. Social Security Administration's names database to a list of the world's Olympic athletes, and both method and data are freely available [4].

Names that did not appear in any of the reference datasets were submitted to Genni [5], a service that takes into account the perceived ethnicity of first and last names to improve estimates of the cultural associations between first names and gender.

Finally, names with posterior probabilities or Genni scores of $\geq 0.9$ and $\leq 0.1$ were associated with the labels *woman* and *man*, respectively. Conservatively, the 18.8% of names with scores between 0.1 and 0.9 were not included in our analyses.

### 1-E. Identifying Field Entries and Exits

The most critical element of the algorithm that identifies field entries and exits is the one that *orders* the affiliations of a given researcher. As noted above, we consider only those affiliations with either a clear academic role or a start date. When all of a researcher's affiliations have a date, ordering is trivial. In fact, because 95.8% of the affiliations had a start date, they were easily ordered in the vast majority of cases. In the remaining cases, when all of a researcher's affiliations are associated with one of the clear academic career stages considered in this paper (see Section 1-A), and a researcher has no more than one affiliation per stage, we assumed an order of bachelor's degree → master's degree → PhD → postdoctoral researcher → professor/department head. In this case, again, ordering is trivial.

The only difficult remaining cases are those researchers whose affiliations are a mixture of dates without career stages and career stages without dates. In this case, we used a simple algorithm that attempts to interleave affiliations. The algorithm takes advantage of any cases where the researcher lists both a date and a career stage, using such affiliations as an anchor to sort the other affiliations by dates and career stages, again under the same ordered career stage assumption as above. Implementations of these algorithms have been made publicly available.

[1] https://github.com/seatgeek/fuzzywuzzy

**Table S4:** Regions represented in the ORCID dataset, alongside the corresponding countries.

| **Region** | **N. of Transitions** | **Countries** |
|---|---|---|
| Europe | 41293 | Albania, Andorra, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Channel Islands, Croatia, Czech Republic, Denmark, Estonia, Faroe Islands, Finland, France, Germany, Gibraltar, Greece, Greenland, Guernsey, Holy See (Vatican City State), Hungary, Iceland, Ireland, Isle of Man, Italy, Jersey, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Montenegro, Netherlands, Norway, Poland, Portugal, Republic of Moldova, Romania, Russian Federation, San Marino, Serbia, Slovakia, Slovenia, Spain, Svalbard and Jan Mayen, Sweden, Switzerland, The Former Yugoslav Republic of Macedonia, Ukraine, United Kingdom of Great Britain and Northern Ireland |
| Northern America | 29070 | Canada, United States of America |
| Asia | 18730 | Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Bhutan, Brunei Darussalam, Cambodia, China, China, Hong Kong Special Administrative Region, China, Macao Special Administrative Region, Cyprus, Democratic People's Republic of Korea, Georgia, India, Indonesia, Iran (Islamic Republic of), Iraq, Israel, Japan, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Lao People's Democratic Republic, Lebanon, Malaysia, Maldives, Mongolia, Myanmar, Nepal, Occupied Palestinian Territory, Oman, Pakistan, Philippines, Qatar, Republic of Korea, Saudi Arabia, Singapore, Sri Lanka, Syrian Arab Republic, Taiwan, Tajikistan, Thailand, Timor-Leste, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Viet Nam, Yemen |
| Latin America and the Caribbean | 17636 | Anguilla, Antigua and Barbuda, Argentina, Aruba, Bahamas, Barbados, Belize, Bolivia (Plurinational State of), Brazil, Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Curacao, Dominica, Dominican Republic, Ecuador, El Salvador, French Guiana, Grenada, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Jamaica, Martinique, Mexico, Montserrat, Netherlands Antilles, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, United States Virgin Islands, Uruguay, Venezuela (Bolivarian Republic of), Virgin Islands (British) |
| Oceania | 3127 | American Samoa, Australia, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Micronesia (Federated States of), Nauru, New Caledonia, New Zealand, Papua New Guinea, Pitcairn Islands, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu, Wallis and Futuna |
| Africa | 2892 | Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Libyan Arab Jamahiriya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Morocco, Mozambique, Namibia, Niger, Nigeria, Réunion, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Togo, Tunisia, Uganda, United Republic of Tanzania, Western Sahara, Zambia, Zimbabwe |

*Note.* Regions are listed in decreasing order of transition counts. Transitions to institutions in Oceania and Africa were not analyzed separately due to insufficient statistical power, but are included for completeness.

## 2. Alternative Theoretical Perspectives on Gender Segregation in Academia

### 2-A. Workload

**Background.** Although academic careers generally demand long hours, fields do differ in their workloads (particularly with respect to on-campus hours; [1]) and thus in the extent to which succeeding in them is compatible with work-life balance. Multiple theoretical traditions suggest this variability should have a differential impact on women's and men's career trajectories. For some scholars, this differential impact is due to *intrapersonal factors* such as women's and men's preferences and choices. More women than men report that they value flexibility in work schedules and achieving some level of work-life balance [6–9], so fields that require long hours—particularly on-campus hours, away from home and family—may lead women (but not men) to opt out. This desire for flexibility is thought to be driven in part by women's decisions around childbearing and -rearing (e.g., [10]). Other scholars emphasize the *structural factors* (e.g., societal expectations) that make it such that women, regardless of their own preferences, have more domestic responsibilities than men do (e.g., [11, 12])—a longstanding inequality recently laid bare by the COVID-19 pandemic (e.g., [13–15]). While these two theoretical perspectives may disagree about the reasons (intrapersonal vs. structural) why women and men would differ in their pursuit of fields that demand long and/or inflexible hours, they converge on the prediction that these fields contribute to gender segregation in academia by creating additional obstacles for women relative to men.

Although this prediction is intuitive, so far the evidence that workload differences between fields contribute to patterns of gender segregation in academia is inconclusive. For instance, while the gender differences in work-life balance preferences (e.g., [6, 8]) and self-reported hours worked per week (e.g., [16]) are well documented, we know of no evidence that relates these gender differences to lower participation rates for women in academic fields with higher (vs. lower) workloads. Moreover, while parenthood imposes a heavier penalty on women's academic careers than on men's, the magnitude of this penalty does not seem to differ substantially across fields (e.g., [17, 18]). The only study to date that systematically investigated the relationship between academic fields' workloads and their gender composition found a modest relationship for on-campus workload specifically, whereby fields with higher on-campus workloads graduated fewer female PhDs, $r = -0.32$, $p = .09$ [1]. However, the magnitude of this relationship was greatly diminished when other field characteristics (e.g., FABs) were held constant, further highlighting the uncertainty around the claim that workload differences between fields can explain the patterns of gender segregation in academia.

**Measurement.** To assess the workload of a field, Leslie, Cimpian, and colleagues [1] asked respondents to indicate the number of hours they worked each week "in your office, lab, classroom, or otherwise on campus" and "off campus (e.g., home, coffee shop, other remote site)." Responses to these face-valid items were elicited on an 8-point scale (with 1 to 7 corresponding to 10-hour increments from 10 hours/week to 70 hours/week, and 8 corresponding to >70 hours/week). On- and off-campus workloads were negatively correlated, $r = -.55$, $p < .001$. Responses to the on- and off-campus workload items were averaged (separately) across respondents in a field, resulting in a set of 30 on-campus and 30 off-campus workload scores (one per field). Because Leslie, Cimpian, and colleagues [1] found field-averaged *on-campus* workload to be particularly predictive of women's underrepresentation across fields, our robustness checks focused on this variable as well. The reliability of the field-averaged on-campus workload variable was high, $ICC(2) = .95$.

The on-campus workload measure's relation to gender gaps (as reported by [1]) speaks to its validity. Responses on this measure also lined up with well-known differences between fields. For example, the three fields with the heaviest on-campus workloads were all lab-based natural sciences (biochemistry, molecular biology, and chemistry), whereas the three fields with the lightest on-campus workloads were all in the humanities (philosophy, English literature, and Middle Eastern studies).

### 2-B. Systemizing vs. Empathizing (and Things vs. People)

**Background.** Fields differ in the extent to which they require "systemizing" (that is, the ability and motivation to analyze a topic or phenomenon as a rule-based system of inputs and outputs) or "empathizing" (that is, the ability and motivation to reason in nuanced ways about people and their mental states; e.g., [19, 20]). This variability may have a differential impact on male and female academics' career trajectories because, according to a prominent theoretical perspective, "males spontaneously systemise to a greater degree than do females" on average and, conversely, "females spontaneously empathise to a greater degree than do males" ([21], p. 248; see also [20, 22]). This argument is related to, and builds on, previous evidence that men report preferring occupations that focus on inanimate objects, which are more amenable to systemizing, whereas women report preferring occupations that deal with people (and living things more generally), who are more amenable to empathizing [23, 24] (for a recent meta-analytic review, see [25]).

In a first test of this argument, Billington and colleagues [19] administered measures of systemizing (e.g., "When I learn a language, I become intrigued by its grammatical rules") and empathizing (e.g., "I can tune into how someone else feels rapidly and intuitively") to 415 college students and found that (1) STEM (vs. humanities) majors scored higher in systemizing and lower in empathizing; that (2) men (vs. women) scored higher in systemizing and lower in empathizing; and, critically, that (3) the gender differences in participation in STEM vs. humanities were to some degree explained by the gender differences in systemizing vs. empathizing.[2] These findings were recently replicated in a sample of over 600,000 individuals by Greenberg and colleagues [20]. In addition, inventories of vocational interests suggest large gender differences ($d$ = 0.93) on the things–people dimension, with men reporting a preference to work with things and women reporting a preference to work with people [25] (but see [26]).

While these findings are suggestive, they fall short of demonstrating that field differences in systemizing vs. empathizing contribute to gender segregation in academia. For instance, fields that differ in their systemizing vs. empathizing requirements might differ in other respects as well (e.g., their FABs), so any systemizing–empathizing (or things–people) differences identified so far between fields could be confounded by other field characteristics. Consistent with this possibility, Leslie, Cimpian, and colleagues [1] found a correlation between a field's emphasis on systemizing vs. empathizing and its gender balance (i.e., fields with a stronger emphasis on systemizing relative to empathizing had fewer female PhDs), but this correlation disappeared when adjusting for other field characteristics. Thus, on the balance, the evidence to date for the systemizing–empathizing account of gender segregation in academia is inconclusive.

**Measurement.** To measure the extent to which a field is perceived to rely on systemizing vs. empathizing [21, 22], Leslie, Cimpian, and colleagues [1] asked respondents to rate the extent to which several processes are "involved in doing scholarly work" in their discipline. Two of the items that followed this prompt were intended to capture systemizing (e.g., "identifying the abstract principles, structures, or rules that underlie the relevant subject matter"; $r$ = .47, $p < .001$) and two were intended to capture empathizing (e.g., "having a refined understanding of human thoughts and feelings"; $r$ = .70, $p < .001$). To maximize validity, the phrasing of these items followed closely the definitions of systemizing and empathizing provided in prior work. For example, Baron-Cohen [21] defined systemizing as "the drive to analyse the variables in a system, to derive the underlying rules that govern the behaviour of a system" (p. 248); the wording of the item above mirrors this definition.

Responses to the two empathizing items were averaged and then subtracted from the average of the two systemizing items. The resulting difference score tracks the extent to which a respondent perceives their field to rely on systemizing more than empathizing. These difference scores were averaged across the respondents in a field, resulting in a set of 30 systemizing–empathizing scores (one per field). This field-averaged measure showed very strong reliability, $ICC(2)$ = .97.

To assess the validity of this measure, we sought to replicate systemizing/empathizing differences previously identified among academic disciplines. In a study of 415 college students, Billington and colleagues [19] found that students who majored in physical sciences showed cognitive profiles that favored systemizing over empathizing, whereas the opposite was true of students who majored in the humanities. Although Leslie, Cimpian, and colleagues [1] measured a field's perceived emphasis on systemizing vs. empathizing rather than individuals' cognitive profiles, it should nevertheless be the case that the physical sciences in their sample will show higher systemizing–empathizing scores than the humanities. We were able to match 17 fields from Leslie, Cimpian, and colleagues' survey to fields represented among the college majors investigated by Billington and colleagues.[3] As in prior work, systemizing–empathizing scores were significantly higher among the physical sciences (e.g., physics; $n$ = 8) than among the humanities (e.g., history; $n$ = 9), $t(15)$ = 3.90, $p$ = .001.

Also relevant to this measure's validity, Leslie, Cimpian, et al. [1] found that a field's systemizing–empathizing score was negatively correlated with the percentage of female PhDs in that field, $r = -0.53$, $p = .003$, meaning that fields with a stronger emphasis on systemizing relative to empathizing had larger gaps favoring men. This pattern is consistent with Baron-Cohen's arguments (e.g., [21, 22]) that the distinction between systemizing and empathizing can help explain gender differences in career outcomes.

The systemizing vs. empathizing distinction builds on older arguments about gender differences in preferences for occupations that focus on *things* vs. *people* (e.g., [23, 24]). Thus, the systemizing–empathizing measure constructed by Leslie, Cimpian, and colleagues [1] may also serve as a measure of the extent to which scholarly work in a field focuses on things vs. people. To assess empirically whether it is legitimate to equate the systemizing–empathizing and things–people dimensions, we recruited a new sample of 21 academics (graduate students, postdoctoral researchers, and faculty; 10 women, 10 men, 1 non-binary) and asked them to rate the 30 fields surveyed by Leslie, Cimpian, and colleagues on the things vs. people dimension. Approximately half of the participants were asked, "To what extent does each discipline

[2] Although these results are described in later publications as showing that systemizing–empathizing scores "mediate sex differences in STEM" ([20], p. 12155), no mediation test was actually reported in Billington et al. [19], nor did Billington and colleagues report whether (and to what extent) the relation between gender and STEM participation was weakened when adjusting for systemizing–empathizing scores.

[3] Billington and colleagues [19] included the following fields among the *physical sciences*: mathematics, physics, physical natural sciences, chemistry, computer science, geology, communications, engineering, manufacturing engineering, chemical engineering, mineral science, material science, astrophysics, astronomy, and geophysics. Their *humanities* fields were classics, languages, drama, education, law, architecture, Anglo-Saxon, Norse and Celtic Studies, philosophy, oriental studies, English, linguistics, theology, history, history and philosophy of science, and history of art and music. Eight fields from Leslie, Cimpian, and colleagues' [1] survey were matched with the list of physical sciences above: astronomy, chemistry, communication studies, computer science, Earth sciences, engineering, mathematics, and physics. Nine fields from Leslie, Cimpian, and colleagues' survey were matched with the list of humanities above: art history, classics, education, English literature, history, linguistics, Middle Eastern studies, philosophy, and Spanish.

below involve thinking about non-human things vs. people?" (1 = *only non-human things* to 7 = *only people*). For the other participants, the order of mention of things vs. people was reversed in the question, and the endpoints of the response scale were flipped as well. Each participant rated all 30 fields. We averaged ratings across participants to generate a single *things–people* score per field, with higher scores indicating a greater focus on things relative to people.

The correlation between the ratings of the 30 fields on the things–people dimension (as assessed with the present sample) and the systemizing–empathizing dimension (as assessed in Leslie, Cimpian, et al. [1]) was $r = .83$, $p < .001$. The strength of this correlation, along with the conceptual similarity of the two measures, suggests they likely tap the same underlying construct (e.g., [27, 28]). Thus, we proceeded on the assumption that the systemizing–empathizing measure is also informative about the people–things dimension that prior work has found to be highly predictive of gender differences in career paths (e.g., [23–25]).

### 2-C. Selectivity

**Background.** Fields differ in their selectivity (e.g., [29]). This variability may have a differential impact on male and female academics' career trajectories because, according to current theories of sexual selection, human males have evolved to be more variable than human females along a range of physical and psychological dimensions (e.g., [30]). That is, even if women and men may not differ *on average* with respect to a certain trait or ability (e.g., intelligence, mathematical ability), differences in *variability* may still exist, with men being overrepresented at both the high and low ends of the relevant distributions (e.g., [31–35]; but see [36, 37]). Thus, the more selective a field is, the more likely it is to recruit individuals from the extreme high end of the relevant ability distributions, and as a result the bigger the gender gaps favoring men should be—because, on this argument, men are increasingly overrepresented relative to women as one approaches the tails. A version of this argument has been invoked specifically as an explanation for gender gaps among STEM faculty (e.g., [10, 30, 38]). For instance, Stewart-Williams and Halsey [30] articulated this explanation as follows: "To the extent that greater male variability results in more males than females occupying the upper echelons of ability, whether for specific aptitudes or general cognitive ability, this may help to explain why more males than females occupy the upper echelons of certain fields in STEM" (p. 13).

Such in-principle arguments are common in the literature (see also the much-discussed speech by Summers [38]). However, empirical evidence that greater male variability, if present, actually translates into higher participation rates for men in fields or positions that are particularly selective is scarce. In fact, the few existing studies relevant to this issue reveal a weak relationship in the *unpredicted* direction, with men being under- rather than overrepresented in fields that are more selective [1, 39, 40]. Thus, even though the claim that selectivity exacerbates gender segregation (specifically, male overrepresentation) is rooted in theory and prominent in academic discourse on this topic, the evidence so far does not support it.

**Measurement.** To assess the selectivity of a field, Leslie, Cimpian, and colleagues [1] asked the following face-valid question of their faculty respondents: "Roughly what percentage of applicants are accepted into your department's PhD program in a typical year?" Responses were elicited on a scale from 1 to 10, with each number corresponding to a 10% increment. Two additional options allowed respondents to indicate that they did not know the answer to this question and that their department did not have a PhD program, respectively. Responses to this item were reverse-scored (so that higher numbers indicate greater selectivity) and then averaged across the respondents in a field, resulting in a set of 30 selectivity scores (one per field). These field-averaged scores scores exhibited strong reliability, $ICC(2) = .82$, meaning that the scores distinguished reliably between fields.

Faculty reported PhD admission rates ranging from less than 10% (e.g., in philosophy) to about 30% (e.g., in chemistry). This range, as well as the relative ordering of the disciplines in terms of admission rates, is consistent with publicly available admission statistics from top, research-intensive U.S. universities—the category of universities that Leslie, Cimpian, and colleagues [1] sampled with their survey.[4] This comparison provides evidence for the validity of Leslie, Cimpian, and colleagues' [1] measure of selectivity.

However, admission rates are only one measure of a field's selectivity. As an alternative operationalization of this construct, we used the average GRE scores of graduate applicants to each field, which reflect the "quality" of each field's pool of applicants [41, 42]. GRE scores were available for all fields except two: linguistics and music theory and composition. Leslie, Cimpian, and colleagues' [1] measure of selectivity was not significantly correlated with the GRE-based measure, $r = .13$, $p = .50$, suggesting that these measures are complementary rather than redundant. Analyses using this GRE-based measure of selectivity as a robustness check replicated the results reported here and in the main text.

[4] For purposes of this validation exercise, we retrieved graduate admissions statistics from three large, research-intensive universities that make these statistics available to the public: University of California, Los Angeles (UCLA); University of Texas at Austin (UT); and University of Michigan, Ann Arbor (UM). These statistics indicated that the survey data are valid: For example, the philosophy faculty surveyed by Leslie, Cimpian, and colleagues [1] reported admission rates lower than 10%, and—in fact—the average admission rate for philosophy PhD programs across UCLA, UT, and UM in Fall 2020 (the latest year available) was 9.1%. At the other end of the spectrum, the chemistry faculty surveyed by Leslie, Cimpian, and colleagues [1] reported admission rates close to 30%, and—consistent with this number—the average admission rate for chemistry PhD programs across UCLA, UT, and UM in Fall 2020 was 29.7%.

**Table S5:** Coefficients (and cluster–robust standard errors) from a conditional logistic regression modeling the probability that an ORCID user **enters** a field (Model 1) and a logistic regression modeling the probability that an ORCID user **exits** a field (Model 2) based on FAB, gender, STEM, and the three alternative explanations

| | **Entering a Field**<br>**Model 1** | **Exiting a Field**<br>**Model 2** |
|---|---|---|
| FAB | $0.118^{***}$ (0.012) | $-0.500^{***}$ (0.016) |
| Workload | $-0.839^{***}$ (0.016) | $-0.563^{***}$ (0.018) |
| Systemizing–Empathizing | $-0.349^{***}$ (0.014) | $0.826^{***}$ (0.024) |
| Selectivity | $-0.903^{***}$ (0.010) | $0.222^{***}$ (0.014) |
| STEM | $1.608^{***}$ (0.023) | $-0.263^{***}$ (0.018) |
| Is Woman | | $0.005$ (0.023) |
| Is Woman × FAB | $-0.961^{***}$ (0.019) | $0.584^{***}$ (0.025) |
| Is Woman × Workload | $0.209^{***}$ (0.029) | $0.083^{**}$ (0.029) |
| Is Woman × Systemizing–Empathizing | $0.354^{***}$ (0.023) | $-0.475^{***}$ (0.038) |
| Is Woman × Selectivity | $0.258^{***}$ (0.016) | $-0.078^{**}$ (0.023) |
| Is Woman × STEM | $-1.202^{***}$ (0.035) | $0.179^{***}$ (0.034) |
| Constant | | $-1.468^{***}$ (0.013) |

*Note.* The coefficients are in log-odds. All continuous predictors were scaled such that *M* = 0 and *SD* = 0.5. The categorical predictor *Is Woman* was coded such that 0 = *man* and 1 = *woman*. Given that the models also include interactions between *Is Woman* and field characteristics, the coefficients for these characteristics in the table above represent their relationships with the dependent variables (e.g., entering a field) among *men* specifically. *STEM* was a categorical predictor that was coded such that 0 = *non-STEM* and 1 = *STEM*. Conditional logit models (such as Model 1) do not estimate an intercept or coefficients for variables that do not vary between "events" (in this case, academics' gender [*Is Woman*]). To assess multicollinearity, which is a potential concern in models with multiple related predictors, we calculated variance inflation factors (VIFs). A general rule of thumb is that VIFs $\leq$ 10 are acceptable [43], although the impact of multicollinearity on estimation accuracy and Type II errors is considerably reduced in large datasets such as ours [44]. Across models, all but three terms had VIFs $<$ 10, and the other three had VIFs that were just above 10. (The largest was 12.08.) Given the size of the ORCID dataset, these values do not provide reason for concern. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

# 3. Robustness to Sampling Bias in the ORCID Data

ORCID usage varies by field, and not all users opt to make their information public. This raises the question of whether the public ORCID dataset can indeed be used to produce unbiased estimates of the relation between FABs and women's and men's career trajectories.

In this section, we describe a set of numerical experiments with synthetic data. In each of these experiments, we first bias the synthetic data in ways that ORCID data might also be biased as a sample of world academics. We then test whether these manipulations detract from our ability to produce unbiased estimates of the "true" relationships between FABs and gender among academics entering and exiting fields. In other words, our goal is to map out the circumstances under which our analyses are robust to potential biases in the ORCID sample, with particular attention to sampling biases by field and gender. All synthetic data and the code used to analyze them are available in the GitHub repository for this paper.

Our numerical experiments correspond to answering whether we can successfully estimate the "true" relationships (i.e., the relationships that are present in the population) with a biased sample under the following conditions:

(1) What if, instead of observing field transitions among the entire population of world academics, we observe only a sample of transitions as large as our ORCID dataset, distributed uniformly across all possible pairs of fields?

(2) What if, due to differences in ORCID usage by field, actual transition counts vary considerably, ranging from 0% to 200% of the uniform values used in the previous experiment?

(3) What if, in addition to heterogeneous transition counts as in the previous experiment, there is also variable bias in how popular ORCID adoption is *by gender*?

(4) What if, in addition to heterogeneous transition counts and variable adoption of ORCID by gender as in the previous experiment, the transitions that are observed are weighted toward non-STEM such that STEM sampling rates are generally lower while non-STEM sampling rates are generally higher? (This experiment reflects a scenario comparable to the observations of Dasler and colleagues in their 2017 study of ORCID usage [45].)

Prior to providing more detail about the various methods for data biasing, we introduce a simple method for creating synthetic data. Our ultimate goal in creating synthetic data is to generate data of the same form that we analyze in our regression analyses of field transitions[5]—that is, counts of women and men observed in transition from field $i$ to field $j$ ($W_{ij}$ and $M_{ij}$, respectively). To that end, let there be 30 academic fields $i = 1, 2, \ldots, 30$, each with a FAB value $x_i$ and a fraction of women $w_i$. Values of $w$ and $x$ are drawn at random. In particular, $x_i \sim$ UNIFORM$[2, 5]$ and $w_i \sim$ UNIFORM$[0.15, 0.85]$, IID. These ranges were meant to reflect ranges observed in empirical data.

Our data generation method proceeds by first stochastically choosing the total number of migrants from $i$ to $j$ of any gender, and then stochastically choosing whether each migrant is a man or a woman depending on a (possibly biased) function of parameters. First, let the total number of people moving from $i$ to $j$ be given by $N_{ij}$, an integer drawn from a geometric distribution with mean $\bar{N}$. Let each of these $N_{ij}$ people be a woman independently of all others with probability $p_{ij}$ and a man otherwise, according to the model

$$p_{ij} = \frac{1}{1 + \exp[-\beta(x_j - x_i) - \log \frac{w_i}{1-w_i} - b_{ij}]} \; , \qquad (1)$$

where $b_ij$ is a gender bias term indicating uneven sampling of women and men transitioning between field $i$ and field $j$.

Note that when $\beta = 0$ and $b_{ij} = 0$, then $p_{ij} = w_i$. In other words, when there is no effect of the FAB variable $x$ and no gender bias in ORCID participation, the gender ratio among migrants from $i$ to $j$ is exactly the gender balance in the field of emigration $w_i$. Having computed values for $p_{ij}$, we then assign a simulated gender to each migrant, resulting in values for $W_{ij}$ and $M_{ij}$.

We now show that, under increasingly extreme biasing of the counts $M$ and $W$ (see Numerical Experiments 1 through 4 below), it is nevertheless possible to recover $\beta$, which measures the effect of the FAB variable $x$ on $M$ and $W$. In other words, we show that the model remains numerically consistent under a range of possible biases in the ORCID data.

Let **Numerical Experiment 1** be a simple test of consistency with a dataset of size equal to the actual number of transitions observed in our ORCID dataset. In plain language, the data for Numerical Experiment 1 shows no bias at all: We have exactly the same number of transitions as the data under study, drawn randomly from the population of interest (i.e., world academics). For true values of $\beta$, we ask whether we are able to accurately estimate $\beta$ using the regression analysis described in the "Taking into Account the FABs of the Source and Destination Fields Simultaneously" section of the main text.

[5] As noted in the main text, we used the modeling strategy that simultaneously takes into account the FABs of the fields that academics exit and enter, which provides the most precise way of assessing how FABs relate to gender differences in career trajectories. This strategy is also simpler, requiring a single model rather than separate models for field entries and exits.

Let **Numerical Experiment 2** be similar to Numerical Experiment 1, in that we target the same number of synthetically generated transitions, but instead of retaining the original target counts $N_{ij}$, we multiply them by a value chosen uniformly at random between $0$ and $2$. In plain language, Numerical Experiment 2 imagines that ORCID users are a variable and noisy sample where transitions between any two fields are observed at a rate chosen randomly between $0\times$ and $2\times$ the average rate.

Let **Numerical Experiment 3** be identical to Numerical Experiment 2 in terms of dataset size and heterogeneity in between-field transition counts, but with a nonzero gender bias $b$. In particular, we draw $b_{ij}$ from a standard normal distribution $N(0, 1)$ IID for each flow $i \to j$. For Numerical Experiment 3, we independently repeat the process ten times. In plain language, Numerical Experiment 3 imagines that ORCID users are a variable and noisy sample where transitions are observed at a rate chosen randomly between $0\times$ and $2\times$ the average, just like Numerical Experiment 3, but with an additional constraint. This experiment assumes that there is gender bias in sampling, such that one gender is more likely to have public-facing ORCID profiles than the other, with said biases drawn differently for each $i \to j$ flow.

Finally, let **Numerical Experiment 4** be identical to Numerical Experiment 3, but with an additional layer of bias: The data are sampled to reflect somewhat lower ORCID usage among those who are currently in, or have ever been in, STEM fields [45]. Whereas in Numerical Experiment 3 transitions were observed at a rate between $0\times$ and $2\times$ the average that was *independent* of the fields $i$ and $j$ involved in the $i \to j$ flow, here we let these rates depend on whether $i$ or $j$ is a STEM field. Specifically, if $i$ or $j$ is a STEM field, rates were chosen uniformly between $0\times$ and $0.5\times$, while if neither $i$ nor $j$ is a STEM field, rates were chosen uniformly between $1.5\times$ and $2\times$.

Across all four experiments, and for a variety of choices of true $\beta$ and repeatedly redrawing $b_{ij}$ for ten technical replicates of Numerical Experiments 3 and 4, we find that the estimated $\beta$ values are a close match to the true $\beta$ values (see Figure S2). In other words, variable gender and field sampling biases do not interfere with the regression's ability to accurately estimate the $\beta$ coefficients of the FAB variable.

Still, there remains the possibility that there are varieties of sampling bias that could affect our results, particularly when this bias is correlated with FABs. For instance, if it is the case that making a transition between fields has a differential relationship with men's and women's choices to make a public-facing ORCID, which is further correlated with or magnified by a field's FAB, the regression models used in this study would not be able to identify and account for this form of bias. However, this bias scenario is unlikely. Under this scenario, for instance, it would have to be the case that more women than men who transition from physics (brilliance-oriented FAB) to psychology (effort-oriented FAB) *just so happen* to have a public-facing ORCID profile, while more men than women who transition from psychology to physics *just so happen* to have a public-facing ORCID profile. Although we cannot rule out this type of "just so" sampling bias, the probability of such systematic coincidences across all 870 possible $i \to j$ pairs of field transitions in this dataset is small.

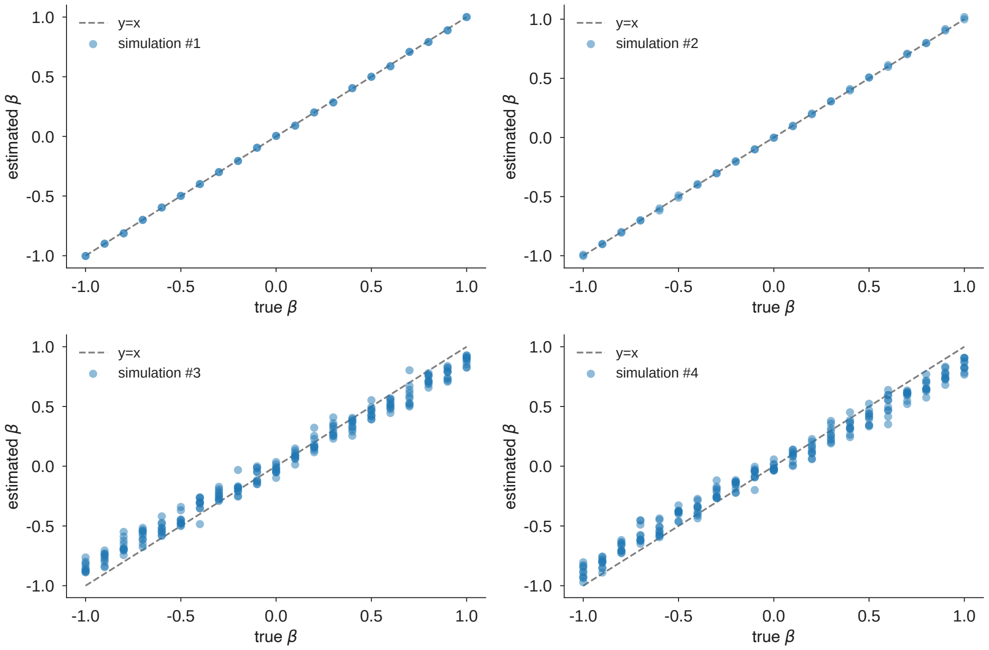


**Figure S2:** Results of simulations for Numerical Experiments 1–4 (left to right, top to bottom). In all cases, the estimated parameters are close to the true parameters.

# References

[1] S.-J. Leslie, A. Cimpian, M. Meyer, and E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, Science **347**, 262 (2015).

[2] ORCID Public Data File Use Policy, https://orcid.org/content/orcid-public-data-file-use-policy (2019).

[3] W. H. Batchelder and A. K. Romney, Test theory without an answer key, Psychometrika **53**, 71 (1988).

[4] I. Van Buskirk, A. Clauset, and D. B. Larremore, An open-source cultural consensus approach to name-based gender classification (2022).

[5] V. I. Torvik 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* (Library of Congress, Washington DC, USA, 2016).

[6] K. O. McCabe, D. Lubinski, and C. P. Benbow, Who shines most among the brightest?: A 25-year longitudinal study of elite STEM graduate students., Journal of Personality and Social Psychology **119**, 390 (2019).

[7] C. Hakim, Women, careers, and work-life preferences, British Journal of Guidance & Counselling **34**, 279 (2006).

[8] K. Ferriman, D. Lubinski, and C. P. Benbow, Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood, Journal of Personality and Social Psychology **97**, 517 (2009).

[9] S. E. Rhoads, *Taking sex differences seriously* (San Francisco, CA: Encounter Books, 2004).

[10] S. J. Ceci and W. M. Williams, Understanding current causes of women's underrepresentation in science, Proceedings of the National Academy of Sciences **108**, 3157 (2011).

[11] A. Hochschild and A. Machung, *The second shift: Working families and the revolution at home* (New York: Penguin, 2012).

[12] J. Williams, *Unbending Gender: Why Family and Work Conflict and What To Do About It* (New York: Oxford University Press, 2001).

[13] B. P. Gabster, K. van Daalen, R. Dhatt, and M. Barry, Challenges for the female academic during the COVID-19 pandemic, The Lancet **395**, 1968 (2020).

[14] T. M. Yildirim and H. Eslen-Ziya, The differential impact of COVID-19 on the work conditions of women and men academics during the lockdown, Gender, Work & Organization **28**, 243 (2021).

[15] S. J. G. Ahn, E. T. Cripe, B. Foucault Welles, S. C. McGregor, K. E. Pearce, N. Usher, and J. Vitak, Academic caregivers on organizational and community resilience in academia (fuck individual resilience), Communication, Culture and Critique **14**, 301 (2021).

[16] D. Lubinski, C. P. Benbow, R. M. Webb, and A. Bleske-Rechek, Tracking exceptional human capital over two decades, Psychological Science **17**, 194 (2006).

[17] E. A. Cech and M. Blair-Loy, The changing career trajectories of new parents in STEM, Proceedings of the National Academy of Sciences **116**, 4182 (2019).

[18] A. C. Morgan, S. F. Way, M. J. Hoefer, D. B. Larremore, M. Galesic, and A. Clauset, The unequal impact of parenthood in academia, Science Advances **7**, eabd1996 (2021).

[19] J. Billington, S. Baron-Cohen, and S. Wheelwright, Cognitive style predicts entry into physical sciences and humanities: Questionnaire and performance tests of empathy and systemizing, Learning and Individual Differences **17**, 260 (2007).

[20] D. M. Greenberg, V. Warrier, C. Allison, and S. Baron-Cohen, Testing the Empathizing–Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people, Proceedings of the National Academy of Sciences **115**, 12152 (2018).

[21] S. Baron-Cohen, The extreme male brain theory of autism, Trends in Cognitive Sciences **6**, 248 (2002).

[22] S. Baron-Cohen, *The essential difference: The truth about the male and female brain* (New York: Basic Books, 2003).

[23] R. Lippa, Gender-related individual differences and the structure of vocational interests: The importance of the people–things dimension., Journal of Personality and Social Psychology **74**, 996 (1998).

[24] D. J. Prediger, Dimensions underlying Holland's hexagon: Missing link between interests and occupations?, Journal of Vocational Behavior **21**, 259 (1982).

[25] R. Su, J. Rounds, and P. I. Armstrong, Men and things, women and people: A meta-analysis of sex differences in interests., Psychological Bulletin **135**, 859 (2009).

[26] V. Valian, Interests, gender, and science, Perspectives on Psychological Science **9**, 225 (2014).

[27] J. A. Shaffer, D. DeGeest, and A. Li, Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs, Organizational Research Methods **19**, 80 (2016).

[28] J. K. Harter and F. L. Schmidt, Conceptual versus empirical distinctions among constructs: Implications for discriminant validity, Industrial and Organizational Psychology **1**, 36 (2008).

[29] H. Okahana, E. Zhou, and J. Gao, *Graduate enrollment and degrees: 2009 to 2019* (Washington, DC: Council of Graduate Schools, 2020).

[30] S. Stewart-Williams and L. G. Halsey, Men, women and STEM: Why the differences and what should be done?, European Journal of Personality **35**, 3 (2021).

[31] L. V. Hedges and A. Nowell, Sex differences in mental test scores, variability, and numbers of high-scoring individuals, Science **269**, 41 (1995).

[32] M. Karwowski, D. M. Jankowska, J. Gralewski, A. Gajda, E. Wiśniewska, and I. Lebuda, Greater male variability in creativity: A latent variables approach, Thinking Skills and Creativity **22**, 159 (2016).

[33] M. C. Makel, J. Wai, K. Peairs, and M. Putallaz, Sex differences in the right tail of cognitive abilities: An update and cross cultural extension, Intelligence **59**, 8 (2016).

[34] A. Baye and C. Monseur, Gender differences in variability and extreme scores in an international context, Large-scale Assessments in Education **4**, 1 (2016).

[35] Y. Xie and K. A. Shauman, *Women in science: Career processes and outcomes* (Cambridge, MA: Harvard University Press, 2003).

[36] A. Feingold, Gender differences in variability in intellectual abilities: A cross-cultural perspective, Sex Roles **30**, 81 (1994).

[37] R. E. O'Dea, M. Lagisz, M. D. Jennions, and S. Nakagawa, Gender differences in individual variation in academic grades fail to fit expected patterns for STEM, Nature Communications **9**, 3777 (2018).

[38] L. H. Summers, Remarks at NBER conference on diversifying the science & engineering workforce, The Office of the President, Harvard University (2005).

[39] C. Iannelli, A. Gamoran, and L. Paterson, Fields of study: Horizontal or vertical differentiation within higher education sectors?, Research in Social Stratification and Mobility **57**, 11 (2018).

[40] M. Hällsten, The structure of educational decision making and consequences for inequality: A Swedish test case, American Journal of Sociology **116**, 806 (2010).

[41] A. Cimpian and S.-J. Leslie, Response to comment on "Expectations of brilliance underlie gender distributions across academic disciplines", Science **349**, 391 (2015).

[42] D. K. Ginther and S. Kahn, Comment on "Expectations of brilliance underlie gender distributions across academic disciplines", Science **349**, 391 (2015).

[43] J. Miles, Tolerance and variance inflation factor, in *Wiley StatsRef: Statistics Reference Online* (John Wiley & Sons, Ltd., 2014).

[44] C. H. Mason and W. D. Perreault, Collinearity, power, and interpretation of multiple regression analysis, Journal of Marketing Research **28**, 268 (1991).

[45] R. Dasler, A. Deane-Pratt, A. Lavasa, L. Rueda, and S. Dallmeier-Tiessen, Study of ORCID adoption across disciplines and locations (2017).