**Degrees of Difference: Gender Segregation of US Doctorates by Field and Program Prestige**

Kim A. Weeden
*Cornell University*

Sarah Thébaud
*University of California, Santa Barbara*

Dafna Gelbgiser
*Cornell University*

# Degrees of Difference: Gender Segregation of U.S. Doctorates by Field and Program Prestige

Kim A. Weeden,[a] Sarah Thébaud,[b] Dafna Gelbgiser[a]

a) Cornell University; b) University of California, Santa Barbara

**Abstract:** Women earn nearly half of doctoral degrees in research fields, yet doctoral education in the United States remains deeply segregated by gender. We argue that in addition to the oft-noted segregation of men and women by field of study, men and women may also be segregated across programs that differ in their prestige. Using data on all doctorates awarded in the United States from 2003 to 2014, field-specific program rankings, and field-level measures of math and verbal skills, we show that (1) "net" field segregation is very high and strongly associated with field-level math skills; (2) "net" prestige segregation is weaker than field segregation but still a nontrivial form of segregation in doctoral education; (3) women are underrepresented among graduates of the highest-and to a lesser extent, the lowest-prestige programs; and (4) the strength and pattern of prestige segregation varies substantially across fields, but little of this variation is associated with field skills.

**Keywords:** gender segregation; prestige segregation; field segregation; gender inequality; higher education; women in STEM

W OMEN earn 60 percent of baccalaureate degrees and 46 percent of doctoral degrees in research fields (National Science Foundation [NSF] 2015a), yet higher education in the United States remains deeply segregated by gender. To date, the literature on educational segregation has focused on the distribution of men and women across fields of study, how this distribution varies over time and space, and its consequences for gender inequality in career outcomes (Charles and Bradley 2002, 2009; England and Li 2006; England et al. 2007; Barone 2011; Bobbitt-Zeher 2007; Mann and DiPrete 2013, 2016; NSF 2015b; Ransom 1990). We extend this line of research by offering a multidimensional analysis of segregation in doctoral education across fields of study and across PhD-granting programs that differ in their prestige.

Our interest in prestige segregation stems from four sources. At the most basic level, prestige segregation, like field segregation, is an indicator of the extent to which men and women's educational outcomes are equal. However, the two types of segregation are conceptually and empirically distinct. Even in a hypothetical world in which every field graduates the same proportions of men and women, men may be overrepresented among degree recipients from the highest ranked programs and women among degree recipients from the lowest ranked programs. In this world, gender integration exists with respect to field, but not with respect to program prestige.

Second, field and prestige segregation represent two qualitatively different forms of gender inequality in higher education. Prestige segregation is inherently vertical, meaning that segregation occurs across categories that are ordered from

high to low. Field segregation, by contrast, is horizontal: the boundaries between fields define qualitatively different positions, but fields "represent distinctions more of kind than of grade" (Charles and Bradley 2002:574). To be sure, some scholars estimate vertical segregation by identifying an external continuous variable that is assumed to capture distinctions of grade among fields (e.g., the average wages of new graduates), applying this variable to fields, and calculating the share of the overall association between gender and fields of study that is captured by this variable (England et al. 2007; Barone 2011). Prestige segregation can be understood as a complementary (and more direct) measure of vertical segregation in higher education.

Third, prestige segregation presages gender inequality in the jobs for which the doctoral degree is a gateway. Prior research has established a strong, positive correlation between the prestige of doctorates' degree-granting institutions and their later career success (Burris 2004; Long and Fox 1995). This correlation may be driven by differences across programs in the talent of incoming students, the quality of the training they receive, the level of financial support they enjoy, the professional networks they develop, the "halo" effect of obtaining a degree from a high-prestige program, or some combination. For our purposes, the causal mechanism is less critical than the correlation: if women are less likely than men to receive their doctoral degrees from high-prestige programs, they will also be underrepresented in the labor market positions in which high-prestige doctorates have a competitive advantage.

Finally, prestige segregation is an expected, albeit underappreciated, outcome of more general social processes identified in the gender inequality and organizational literatures. We focus on five such processes: (1) sorting based on gender differences in readily observed indicators of ability (and their unobserved correlates), (2) sorting based on gender-biased self-assessments of ability, (3) self-selection based on gender-differentiated preferences for different program attributes that are correlated with prestige, (4) prestige-linked organizational strategies surrounding admissions, and (5) gender-specific attrition from graduate programs. The first two processes are often deployed to understand the sources of field segregation and in particular women's underrepresentation in scientific (STEM) fields. The third and fourth processes, which focus on the attributes of the degree-granting programs themselves, are rarely discussed in the educational segregation literature, presumably because the organizations that represent fields (e.g., scholarly societies) have little input into admissions and training. The fifth process, gender-specific attrition, is often analyzed as an outcome in its own right but rarely linked to segregation. Our theoretical contribution is to articulate possible sources of prestige segregation, which we accomplish by drawing from a variety of literatures that are rarely in dialogue with each other.

Our main contribution, however, is empirical: we offer a systematic analysis of levels and patterns of field and prestige segregation among research doctoral degree recipients in the United States.[1] This analysis is based on a national census of earned doctoral degrees from the Integrated Postsecondary Education Data System (IPEDS), which we merged to prestige rankings of doctoral programs from the National Academies of Sciences (NAS) and to field-level measures of math and verbal skills

drawn from the General Record Exam (GRE) (Educational Testing Service [ETS] 2008). We analyze these data with log-multiplicative association models that allow us to tease out levels and patterns of field and prestige segregation, estimate cross-field variations in prestige segregation, and quantify the extent to the which field segregation and cross-field variations in prestige segregation map onto field math and verbal skills.

## Prior Research on Prestige Segregation

Extant research on prestige segregation in higher education is both sparse and inconclusive. In a study of prestige segregation at the undergraduate level, Davies and Guppy (1997) show that men are more likely than women to graduate from selective institutions (as measured by average SAT scores), in lucrative fields (as measured by the average pay of graduates), and in lucrative fields within selective institutions. By contrast, Jacobs (1995, 1999) found few gender differences in the prestige of baccalaureate institutions after adjusting for women's lower representation in STEM fields and the high concentration of STEM fields in high-prestige universities (e.g., Cal Tech, MIT). Quite aside from the disparity in their core findings, studies of undergraduates don't necessarily generalize to doctoral education, for which admissions and training decisions take place at the department level, the academic orientation of the "average" student is greater, and postmatriculation field- or program-switching is uncommon.

Research on prestige segregation among doctoral students is even less well developed, and most of it is very dated. The few studies that exist are inconclusive, in our view, because they deploy data on a very limited range of fields or institutions, conflate the prestige of the doctoral program with the prestige of the university in which it resides, use methods that cannot differentiate prestige segregation from field segregation, or assume a linear relationship between program prestige and the gender composition of graduating cohorts (Fox 1995; Gilford and Snyder 1977). We will take up these data and methodological issues below. First, however, we motivate our analysis by identifying social processes that might plausibly generate prestige segregation and the empirical patterns of segregation that these processes imply.

## Sources of Prestige Segregation

The gender composition of a cohort of doctorates is a function of the gender composition of incoming cohorts and gender-specific attrition. We will first discuss social processes that could generate segregation at the point of creating a cohort, then gender-specific attrition. Throughout, we will devote more attention to prestige segregation than to field segregation, given that the theoretical literature on field segregation is already quite extensive.

## Self-Selection Based on Measured Ability

To begin, we assume that applications to and rejection from grad school is costly, that students will only apply to programs for which they believe they have a nonzero probability of admission, and that they will only matriculate at programs for which they believe they have a nonzero probability of completion. We also assume that their assessments of these chances are affected by their prior academic performance and their beliefs about whether they will be competitive. This, in turn, is plausibly associated with program prestige: students expect the competition for slots in a cohort and for faculty time (and other resources) to be stiffer at higher-ranked programs than at lower-ranked programs. As a result, students who are the most able (or who perceive that they are the most able) will be overrepresented among applicants to (and matriculates of) high-prestige programs.

The "measured ability" variant of the selection argument can only help us understand gender segregation in doctoral education if observed indicators of academic ability and the unobserved indicators with which they are correlated differ systematically by gender. In this regard, studies of high school students typically find modest gender differences in test scores at the mean but nontrivial gender differences in the distributions: on tests of math and scientific reasoning, young men outnumber women at the top of the distribution but also at the bottom (i.e., greater male variance); on tests of verbal reasoning and writing ability, young women outnumber young men at the top of the distribution, although to a lesser extent (Penner and Paret 2008; Riegle-Crumb et al. 2012; Makel et al. 2016). Among GRE test takers in research fields, men's average and 75th percentile scores are approximately 20 points higher than women's average and 75th percentile scores on the verbal test and a much more substantial 60 points higher than women's average and 75th percentile scores on the quantitative test (ETS 2008).[2] Studies of gifted students show even more extreme gender gaps (favoring men) at the right tail of the math test score distributions and smaller gender gaps (favoring women) at the right tail of verbal test score distributions, although the latter are less stable across different tests of ability and prior achievement (Makel et al. 2016).

At the undergraduate level, gender differences in standardized math test scores and prior math preparation account for only a modest portion of the gender gap in STEM major choice (Mann and DiPrete 2013; Morgan, Gelbgiser, and Weeden 2013; Riegle-Crumb et al. 2012; Xie and Shauman 2003), in part because men are overrepresented in the lower tails of the distributions as well as the upper tails (i.e., greater male variance). At the graduate level, however, the pool of applicants is presumably drawn from the upper tails of the distributions of test scores and academic preparation, where gender differences in the easily observed indicators of ability are more substantial. Assuming potential applicants use these scores and their correlates to assess their likelihood of success, it follows that the level of math skills associated with a given field will be positively associated with male overrepresentation in that field. The level of verbal skills will have much weaker association with female overrepresentation, given the less extreme gender gaps in the right tail of observed indicators of verbal ability.

Could these gender differences in measured ability at the top of the distribution also generate prestige segregation? If applicants with relatively modest

qualifications—in GRE scores or in the measures of ability and achievement that are correlated with GRE scores—select out of the applicant pools at the highest-prestige programs, and if women are overrepresented among those with modest (measured) qualifications, it follows that women will be underrepresented in the highest prestige programs and overrepresented in lower-prestige programs. Furthermore, levels of prestige segregation are likely to be stronger in math-intensive fields than in verbal-intensive fields, again because of the much smaller gender gaps in the right tail of verbal ability.

## Self-Selection Based on Perceived Ability

Segregation in higher education can also result from gender differences in perceived or self-assessed ability. A now voluminous body of experimental and survey research shows that men and women believe that men are generally more competent and capable than women and that this gender gap in expectations of others' competence is especially strong when the task is associated with stereotypically male traits and abilities such as math reasoning or higher-order cognitive thinking (Wagner and Berger 1997; Ridgeway 1997). Gender status beliefs can also inform individuals' self-assessments of competence at career-relevant tasks: women evaluate their own competence and abilities more negatively than men with the same measured ability, which in turn affects their career-relevant educational decisions (Correll 2001, 2004; Foschi 2008). Consistent with this argument, gender differences in self-assessed math ability have been shown to affect gender differences in STEM (baccalaureate) major aspirations and choices and, at the aggregate level, field segregation (Correll 2001, 2004; Mann and DiPrete 2016).

The logic of the self-assessed ability argument implies prestige segregation as well as field segregation. If, on average, women underestimate their competence relative to men with similar ability, women are less likely to believe that they will be competitive for slots at the highest-ranked programs and consequently less likely to apply to these programs. They may also be less likely to matriculate at a highly ranked program, if low self-assessed ability leads them to second-guess the admissions committee's positive decision, through what psychologists have dubbed the "imposter phenomenon" (Clance and O'Toole 1988).[3] The end result, according to this logic, is prestige segregation. Moreover, this prestige segregation will be greatest in math-intensive fields, given the strong cultural beliefs about men's greater competence in math and science. However, gender differences in self-assessed ability also predict women's underrepresentation in the highest prestige programs in "verbal" fields, given beliefs about men's superior general competence in tasks that require higher cognitive ability.

## Self-Selection Based on Program Attributes

In choosing where to apply and matriculate, prospective students presumably also consider a host of program-related factors: distribution of subfields, intellectual "fit" with the target program, availability of mentors and faculty advisors, funding, proximity to family and friends, geographic region, and so forth. Many of these factors are correlated with program prestige. For example, higher-prestige programs

tend to have better funding packages, fewer women or underrepresented minority faculty, and more limited geographic dispersion (NSF 2015a; Trower and Chait 2002).[4]

The key question for prestige segregation is whether men and women give these factors different weight in selecting a graduate program. The available evidence is mixed. On one hand, women are more likely to have familial constraints that lead them to give more weight to geographic location than men (Blau and Ferber 1982; on undergraduates, see Jacobs 1999; on elite faculty, see Schiebinger, Henderson, and Gilmartin 2008). One implication is that women will be more likely to choose a lower-ranked program that is proximate to home or family over a higher-ranked program some distance away; all else equal, and at the aggregate level, this will likely lead to male overrepresentation in high-prestige programs. On the other hand, familial constraints may lead women to attach more weight to the availability of funding (see Berg and Ferber 1983; but see Dwyer, Hodson, and McCloud 2013), which could increase their concentration in high-prestige programs where, on average, funding is likely to be better. And, finally, male and female graduate students have grown more similar to each other on many demographic attributes (Long 2001), and norms of shared parenting have diffused. These changes may equalize gender disparities in the weight that men and women give to geographic proximity to family or funding and in the process reduce the potential for gender-specific preferences to impact segregation patterns.

Men and women may also differ in the types of programs that they select out of a preference for demographic and subfield matching. Demographic matching occurs when students either select into or are more likely to complete programs for which there are faculty mentors of the same gender (Rosser 2004; Wallace and Haines 2004), implying that the segregation of male and female faculty across programs will tend to generate similar patterns of segregation of students across programs. Consistent with this claim, a recent study of 20 economics departments shows that the higher the share of female faculty, the higher the share of female students graduating six years later (Hale and Regev 2014). Similarly, some disciplines (e.g., sociology) have substantial gender segregation by subfield, and "male"- and "female"-dominated subfields are not evenly distributed across high- and low-prestige programs. This form of segregation, too, may generate gender differences in the applicant pools to programs of higher or lower prestige.

As this discussion suggests, gender differences in the weight given to different program attributes can create prestige segregation simply by virtue of the association between these (nonprestige) attributes and program ranking. However, the implication for prestige segregation is not clear, in part because gender differences in preferences can have offsetting implications for aggregate patterns. We simply note that program-linked attributes, when coupled with gender differences in weighting of particular attributes, may generate prestige segregation.

## Prestige-Linked Admissions Decisions

Students' application and matriculation decisions are just one side of the matching process, and doctoral programs' admissions and training decisions will also

affect the gender composition of incoming cohorts and attrition from graduate school. We assume that with the exception of a few noncompetitive programs that admit all applicants who can pay the fees, most doctoral programs must limit the number of students they admit because of limited resources, including faculty time. Admissions committees select applicants who they believe will maximize their "returns," including skilled research or teaching assistance, the creation of new knowledge, the prestige that accrues to a program when a graduate student secures a high-status job, and other contributions to local or institutional goals. Departments also compete with each other to attract "the best" students, however this is defined (Posselt 2016).

How might programs' admissions decisions contribute to prestige segregation? First, the highest-prestige programs may simply extend offers of admission to students (male or female) who are at the very top of the applicant pool with respect to easily observable indicators of ability and academic achievement, under the assumption that they will win the interdepartmental competition for these students often enough to fill their cohorts. Middle- and lower-ranked programs, by contrast, may put more effort into identifying "diamond-in-the-rough" students who would not necessarily draw the attention of top-ranked programs. Prestige segregation can thus result from "gender blind" admissions processes by virtue of uneven gender distributions on the easily observed indicators of ability and achievement. This argument also implies that prestige segregation will be greater in math-intensive fields (see above), given the greater gender gaps in performance on easily observed indicators of ability in math than in verbal skills.

Second, members of admissions committees may hold male-advantaging status beliefs and judge female applicants by stricter standards than male applicants (Milkman, Akinola, and Chugh 2012; Moss-Racusin et al. 2012; Posselt 2016). These biases in admissions are most obviously relevant to field segregation. However, they may also generate prestige segregation if high- and low-ranked programs systematically differ in the strength of faculty members' gender status beliefs or in the extent to which the admissions process creates the conditions for gender beliefs to become salient, such as severe time constraints, premature ranking of candidates, or failing to read past letters of recommendation. In addition, high- and low-prestige programs may differ in the availability of local examples that counteract general cultural beliefs about men's greater competence at higher cognitive tasks in general and math-related tasks in particular.

Third, doctoral programs may make different admissions decisions depending on their position in local prestige orders. According to the theory of middle status conformity ([MSC]; Phillips and Zuckerman 2001), organizations' likelihood of innovating differs according to whether they are high, middle, or low status[5]: High-status and low-status actors have more leeway to innovate because their actions will have little effect on their status, whereas middle-status actors have less leeway to innovate (i.e., more likely to conform) because they are at greater risk of falling in the status order. To be applicable to the graduate admissions context, MSC requires several crucial assumptions: doctoral programs must be aware of their position in the status order and seek to improve their position or at least avoid falling in position; programs' admissions decisions must affect, even if only in the

long term, their position; and it must be possible to identify "conforming" and "nonconforming" admissions decisions.

We argue that in contemporary higher education in the United States, there are strong institutional pressures to diversify the academe along gender (and racial) lines and that "conformity" thus means some measure of adherence to these egalitarian pressures. Over the last two decades, universities and external agencies (e.g., NSF's ADVANCE program) have made large financial investments in diversity programming and infrastructure, and egalitarian discourse is now common within universities, including among administrators who allocate resources. To be sure, the carrots and sticks associated with efforts to increase diversity vary in effectiveness, and one or two cohorts that lack diversity may garner much attention from internal or external observers. In the long term, however, departments can gain a reputation of being "woman friendly" or "woman unfriendly," which can affect the internal allocation of resources and, in the long term, rankings. Indeed, many scholarly societies (e.g., Sociologists for Women in Society, American Chemical Society) explicitly rank programs by their gender diversity, and the new National Research Council (NRC) ranking system includes measures of demographic diversity as component factors in the summary scores (NRC 2010).

According to the logic of MSC, middle-prestige programs are the most likely to be harmed by acquiring a reputation of "woman unfriendliness" and hence are the most likely to consider gender diversity during the admissions process. Low-prestige programs are less likely to conform, according to this theory, because they have little status to lose (Phillips and Zuckerman 2001). High-prestige programs are also less likely to conform to institutional pressures to diversify because they can parlay their high prestige into relative autonomy from the dictates of central administration and into relatively stable pools of highly talented applicants. Consistent with this claim, Posselt (2016) found that admissions committees at elite departments rarely discuss gender diversity, although of course without data on less elite departments, this evidence is only suggestive.

If MSC is on the mark, the relationship between program prestige and gender segregation will be an inverted U shape: women will be underrepresented in high-status programs, overrepresented in middle-status programs, and underrepresented in low-status programs. We might also anticipate the curvilinear pattern will be more pronounced in math-intensive fields, simply because institutional pressures to increase gender diversity are stronger in the STEM fields.

## Gender-Specific Attrition

Even in a hypothetical world of no gender segregation among incoming cohorts, gender-specific attrition rates can produce gender segregation among cohorts of earned doctorates. Studies of attrition consistently show that women are less likely to complete graduate school than men but also that these patterns vary greatly by field and by institution. Within-institution studies find that most of the gender gap in attrition disappears when one adjusts for men's overrepresentation in math and sciences and women's overrepresentation in fields for which an MA is a meaningful terminal degree (Ehrenberg and Mavros 1995; Lott, Gardner, and Powers 2009;

Zwick 1991). Within-field studies, which are more relevant to prestige segregation, show that gender gaps in attrition have narrowed in the past three decades and in many fields are either trivial or nonexistent (Baker 1997; Nettles and Millett 2006; Ampaw and Jaeger 2012).

Is there reason to believe that residual within-field gender differences in attrition vary with program prestige? In this regard, some scholars have argued that women do not perform as well as men in highly competitive, mixed-sex environments (Gneezy, Niederle, and Rustichini 2003), which might make them more likely than men to drop out of high-prestige programs. Others argue that men have better alternatives in the labor market than women and hence pay a lower price for dropping out of graduate school or are less willing to carry debt to complete graduate school (Dwyer et al. 2013). Although neither argument specifically discusses graduate education, their logic implies male overrepresentation in high-prestige programs. On the other hand, women may be more likely to drop out of low-prestige programs than men because of their greater sensitivity to funding (see above), which would have the opposite impact on aggregate patterns of prestige segregation. We don't take a position on which of these countervailing effects will dominate, but merely note that in addition to processes at the point of admissions and matriculation, gender-differentiated attrition could, in theory, generate prestige segregation.

We do not claim to have identified all of the sociological processes that could, in theory, generate prestige segregation, nor will our data allow us to evaluate which processes are at work. The preceding discussion does imply, however, four empirically testable descriptive claims: (1) field segregation will be strongly associated with field-level math skills and moderately associated with verbal skills; (2) independent of field segregation, men and women will be unevenly distributed across programs grouped by their prestige; (3) men will be overrepresented in the highest prestige programs and, as predicted by MSC, in the lowest-prestige programs; and (4) prestige segregation will be stronger in math-intensive fields than in verbal fields, as predicted by self-selection based on "objective" measures of ability or perceived ability.

## Data

We assess levels and patterns of field and prestige segregation using program-level data on earned doctorates from the IPEDS, which covers all colleges and universities that participate in federal financial assistance programs and some that volunteer data. We pool the IPEDS data collected between 2003 and 2014 in order to have sufficient cases to disaggregate fields and prestige groups.[6] An IPEDS crosswalk allows us to reconcile the Classification of Instructional Programs (CIP) scheme used to classify fields in the post-2010 data with the CIP scheme from 2003 to 2010.

We construct measures of program prestige from the 1995 NRC ratings of doctoral programs' faculty quality, which cover 41 research fields, 274 universities, and 3,271 degree-granting programs in universities that report to IPEDS.[7] We chose the 1995 NRC rankings over the alternatives because of its timing, construction,

and coverage. First, the 1995 rankings were published before the cohorts in our data were admitted to graduate school, so logically they could affect these cohorts' admissions decisions. Second, the 1995 rankings are based solely on subjective assessments, whereas the later 2010 NRC rankings are based on a combination of subjective and objective factors, including the program's demographic composition. And, third, unlike the *U.S. News & World Report*'s rankings, the 1995 NRC rankings cover more fields and use consistent evaluation methods across fields.

The NRC, like other alternatives, does not rank all PhD-granting programs. The unranked programs are a heterogeneous mix of programs located in research universities, teaching- or clinical practice–oriented universities, and specialty institutions (e.g., military schools, theological seminaries). To gain leverage on this heterogeneity within unranked programs, we categorize them by the Carnegie type of the university in which they are located: very high research university (RU/VH); high research university (RU/H); and doctoral research university, specialty, and other (DRU/S/O). Although in theory one could estimate segregation across Carnegie type—or even university prestige—in addition to program prestige, in practice this is intractable because of the very strong association between university type and program ranking. Instead, we treat Carnegie type as a partial table relevant only for unranked programs.

For each doctoral program, the 1995 NRC provides an average score on a five-point, Likert-type assessment of faculty quality collected from representatives of other programs in the same field. From these average scores, we create a measure of absolute prestige, which is the rank order of a given program in its field, and a measure of relative prestige, which is the program's position in the percentile distribution of scores in its field. The absolute prestige rankings capture socially meaningful distinctions (e.g., "4th ranked program"). The relative prestige ranking adjusts for the fact that the NRC rates more programs in some fields than in others and assumes that the social meaning of a "top 25" program may be quite different in a field with 185 ranked programs than with just 26 ranked programs.

We categorize the absolute and relative rankings into "prestige buckets," which allows us to increase cell counts, easily incorporate the unranked programs (by treating them as additional categories), and identify nonlinearities in the pattern of prestige segregation. We create 13 "prestige buckets" from the relative prestige measure: 10 that correspond to NRC ratings deciles, and three that correspond to Carnegie types of unranked programs. Similarly, we create 10 buckets from the absolute prestige measure: top 5, 6–10, 11–15, 16–20, 21–25, 26–30, 31–185, and the same three Carnegie types for unranked programs. The 7th absolute prestige bucket, corresponding to programs ranked 31st–185th, cannot be divided further without creating structural zeros because of the limited number of NRC-ranked programs. As it is, there are no programs in Classics and Oceanography ranked 31st–185th and no unranked programs in Comparative Literature or Aerospace Engineering in "DRU/special/other" universities. Rather than assign an arbitrary constant to these zero cells, we calculate the average gender ratio for the field, assign one of the gender-specific cells the value of 1 (e.g., men in Classics in the 31st–185th absolute ranking "bucket"), and assign the other gender-specific cell a

count that will preserve the field-specific gender ratio. Analyses that weight these cells out of the data (not shown) generated nearly identical results.

We measure the math and verbal skills of the 41 NRC fields with the average math and verbal GRE scores of test-takers intending to go to graduate school in a given field of study (ETS 2008). From these GRE scores, we construct two categorical variables, each of which differentiates five skill groups: more than 1 standard deviation below the cross-field mean, between 0.5 and 1 standard deviation below the mean, within a 0.5 standard deviation on either side of the mean, between 0.5 and 1 standard deviation above the mean, and more than 1 standard deviation above the mean.[8] Table S1 in the online supplement lists the 41 NRC fields, their math and verbal GRE scores, and their math and verbal skill groups.

We then match the 3,271 programs ranked in the NRC to the 5,132 programs listed in the IPEDS database as granting doctoral degrees in the NRC research fields. Of the NRC-ranked programs, we could easily match 2,877, or 88 percent, to the IPEDS using the university name and CIP codes aggregated to the level of the 41 NRC fields. In some cases, universities reported doctorates using a less detailed CIP code than the NRC fields: for example, "Romance Studies" rather than "Spanish" and "French." Using lists of graduate students on these department's web sites, we estimate the (current) share of graduate students in that program in Spanish and in French and divide up the counts in the aggregate CIP code that is reported in IPEDs into the two NRC fields accordingly.

We are left with 284 NRC-ranked programs that we cannot easily match to the IPEDS, most of which are in the biological sciences. For 219 of these programs, we estimate the number of graduates in the NRC-ranked program by identifying the CIP code in the IPEDs in which those graduate students are most likely to be reported. For example, if Cellular Biology is a ranked program at University X, and University X didn't report any graduates in "Cellular Biology" but did report graduates in "Biological Sciences, General" and the other biology subfields, we attribute the counts in the general code to "Cellular Biology." For 65 of the unmatched NRC programs, we cannot identify which CIP code is likely to contain the "missing" graduates, so we exclude these programs from the analysis. As a robustness check (not shown), we also ran analyses on data that exclude all 284 unmatched programs and found no noteworthy differences in the results.

Compared to matched programs, unmatched programs tend to have lower NRC ratings (i.e., lower subjective quality) and to reside in universities that are private, do not grant medical degrees, and have comparatively few NRC-ranked fields (see online supplement Table S2). We cannot, of course, know the number or gender composition of graduates from the 65 unmatched and excluded programs, but they constitute a mere 2 percent of the NRC-ranked programs. Their gender composition would need to be wildly different from the ranked programs for their exclusion to have an appreciable impact on the results, and we have no reason to believe this is the case.

Finally, we create cross-classified arrays of field, program prestige (including the unranked categories), and gender from the merged NRC–IPEDs data. The cell counts in these arrays are the number of doctoral degrees: 406,721 in the relative prestige array and 406,726 in the absolute prestige array, where the difference

**Table 1:** Summary measures of gender segregation of U.S. doctorates by field and program prestige.

|  | | *D* | *A* |
|---|---|---|---|
| Field of study (41 fields) | | 0.33 | 2.12 |
| Relative prestige | | | |
| | Percentiles (1–100; unranked) | 0.11 | 1.29 |
| | Percentiles + unranked (1, ..., 100; RU/VH; RU/H; DRU/Other/Special) | 0.12 | 1.29 |
| | "Buckets" (10 deciles; RU/VH; RU/H; DRU/Other/Special) | 0.09 | 1.26 |
| Absolute prestige | | | |
| | Absolute prestige (1, 2, ..., n, unranked) | 0.13 | 1.64 |
| | Absolute prestige (1, 2, ..., n, RU/VH; RU/H; DRU/Other/Special) | 0.14 | 1.64 |
| | "Buckets" (1–5, 6–10 ..., 31–185, RU/VH; RU/H; DRU/Other/Special) | 0.09 | 1.30 |

NOTES: Data are from the IPEDs, 2003–2014. Relative prestige N = 406,721; Absolute prestige N = 406,726. The NRC ranks between 26 and 185 programs, depending on the field.

emerges because of imputation of additional structural zeros in the absolute prestige arrays. These doctorates are distributed across a maximum of 1,066 cells, yielding a minimum average cell count of 381 doctorates. Even with this large sample, we had one empty gender-specific cell in an unranked program, to which we assigned a value that preserves the field-specific gender ratio (see above).

## Summary Measures of Segregation

Table 1 presents two indices of segregation calculated for fields and program prestige, each collapsing across the other dimension. The index of dissimilarity (D) measures the percentage of men or women who would need to change fields (or programs) in order for each field (or program) to have equal percentages of men and women. The log-linear index (A) is the deviation of the field (program) gender ratios from the mean and can be interpreted as the multiplicative factor by which women (or men) are overrepresented in the average field or program (Charles and Grusky 1995).

Table 1 shows, unsurprisingly, that field segregation in doctoral education is extensive. Women (or men) are overrepresented in the average field by a factor of 2.12, and a third of male or female doctorates would need to change fields in order for men and women to receive the same percentage of PhD degrees in all fields (D = 0.33). This estimate of D is slightly lower than the 35–39 percent estimate provided by England and her colleagues (2007) for the early 2000s, most likely because they measure segregation across 202 fields, and D typically increases with the specificity of the units.

Prestige segregation in U.S. doctoral education over this period is weaker but, we argue, still substantively significant. If we group programs by their prestige percentile and differentiate Carnegie groups for unranked programs, we find that 12 percent (D = 0.12) of male (or female) doctorates would need to change programs to achieve integration, and men (or women) are overrepresented by a factor of 1.29 (Table 1). The comparable index values for the absolute prestige rankings are a

bit higher, with a dissimilarity index of 14 percent and a log-linear index of 1.64. These indices shrink when we aggregate programs into larger prestige buckets: D decreases to 0.09 (relative and absolute prestige), and A decreases to 1.26 (relative) or 1.30 (absolute). Even at this more aggregate level, however, the indices indicate nontrivial levels of prestige segregation.

## Log-Linear Models of Segregation

We investigate net levels and patterns of segregation with a series of log-linear and log-multiplicative models developed for gender segregation research (Charles and Grusky 1995, 2004; Weeden and Sørensen 2004:254–257). Because a formal presentation of these models is available elsewhere, we will focus on their conceptual logic and introduce each model in the context of the substantive question it addresses.

### How Much Field and Prestige Segregation Is There?

Model 1, a model of conditional independence, provides a baseline estimate of the total gender-related association in the data. This model fits main effects of gender, field, and prestige and the interaction of prestige and field but does not allow for prestige or field segregation. Not surprisingly, it fails to fit either the relative (columns 1–3) or absolute (columns 4–6) prestige arrays, indicating substantial gender-related association in the data (see Table 2).

The next two models are scaled association models that assume just one form of segregation, without assuming segregation on the other dimensions: Model 2 assumes field segregation but not prestige segregation, and model 3 assumes prestige segregation but not field segregation. Both models improve fit relative to conditional independence. Model 2 (field segregation) reduces the log-likelihood ($L^2$) by 63,356 with the expenditure of 40 degrees of freedom (df) and accounts for 96 percent (relative prestige array) to 97 percent (relative prestige array) of the residual association under conditional independence (see model contrast 1, Table 2). The Bayesian Information Coefficient (BIC) becomes negative, where smaller values indicate better fit. Applied to the relative prestige array, model 3 (prestige segregation) reduces $L^2$ by 6,426 (12 df) and accounts for 9.7 percent of the residual association; fit to the absolute prestige array, the reduction in $L^2$ is 6,229 (10 df), corresponding to a 9.5 percent decline in the test statistic of model 1 (see model contrast 2, Table 2).

Model 4 is a scaled association model that allows for field and prestige segregation simultaneously and estimates prestige segregation scale values (for example) that are purged of the field by gender and field by prestige associations.[9] A negative prestige scale value for top 10th percentile programs would indicate that men are overrepresented in top 10th percentile programs, adjusting for the overrepresentation of men in engineering and the comparatively large size of elite engineering programs compared to elite programs in fields in which women are overrepresented. As we show in Table 2, model 4 reduces $L^2$ by 171 (12 df) in the relative prestige array and by 109 points (9 df) in the absolute prestige array compared to the field segregation model (see model contrast 3, Table 2). For the relative prestige array,

**Table 2:** Fit statistics for basic log-multiplicative models of field and prestige segregation.

| | | Relative Prestige (10 Decile Groups + 3 Unranked) | | | | Absolute Prestige (7 Prestige Buckets + 3 Unranked) | | | |
|---|---|---|---|---|---|---|---|---|---|
| **A.** | **Models** | **L2** | **d.f.** | **Δ** | **BIC** | **L2** | **d.f.** | **Δ** | **BIC** |
| 1 | Conditional independence | 66,086 | 532 | 16.6 | 59,215 | 65,602 | 409 | 16.6 | 60,319 |
| 2 | Field segregation | 2,729 | 492 | 2.7 | −3,625 | 2,245 | 369 | 2.2 | −2,521 |
| 3 | Prestige segregation | 59,660 | 520 | 15.8 | 52,944 | 59,373 | 400 | 16.5 | 54,206 |
| 4 | Field and prestige segregation | 2,559 | 480 | 2.5 | −3,641 | 2,136 | 360 | 2.1 | −2,514 |
| 4* | Model 4, equivalence constraint in unranked programs | 2,572 | 482 | 2.5 | −3,653 | 2,151 | 362 | 2.1 | −2,524 |
| 5a | Math skill groups and prestige segregation | 21,124 | 516 | 8.9 | 14,460 | 20,588 | 396 | 8.8 | 15,473 |
| 5b | Verbal skill groups and prestige segregation | 48,609 | 516 | 14.0 | 41,944 | 48,510 | 396 | 14.0 | 43,395 |

| | | **Change in L2** | **Change in d.f.** | **Explained variability (%)** | **Change in BIC** | **Change in L2** | **Change in d.f.** | **Explained variability (%)** | **Change in BIC** |
|---|---|---|---|---|---|---|---|---|---|
| **B.** | **Model Contrasts** | | | | | | | | |
| 1 | Field segregation (model 2 vs. model 1) | 63,356 | 40 | 95.9 | −62,840 | 63,356 | 40 | 96.6 | −62,840 |
| 2 | Prestige segregation (model 3 vs. model 1) | 6,426 | 12 | 9.7 | −6,271 | 6,229 | 9 | 9.5 | −6,113 |
| 3 | Net prestige segregation (model 4 vs. model 2) | 171 | 12 | 6.2 | −16 | 109 | 9 | 4.9 | 7 |
| 4 | Net field segregation (model 4 vs. model 3) | 57,101 | 40 | 95.7 | −56,585 | 57,237 | 40 | 96.4 | −56,720 |
| 5 | Constant segregation in unranked programs (models 4* vs. model 4) | 13 | 2 | 0.5 | −12 | 15 | 2 | 0.7 | −11 |
| 6 | Math skill segregation (model 5a vs. model 3) | 38,536 | 4 | 64.6 | −38,485 | 38,785 | 4 | 65.3 | −38,733 |
| 7 | Verbal skill segregation (model 5b vs. model 3) | 11,052 | 4 | 18.5 | −11,000 | 10,863 | 4 | 18.3 | −10,811 |

Δ measures the percentage of cases misclassified under the relevant model.
NOTES: Data are from the IPEDs, 2003–2014. Relative prestige N = 406,721; Absolute prestige N = 406,726.

the value of BIC is smaller for model 4 than for model 2, meaning that model 4 is preferred by this test of model fit. For the absolute prestige array, however, BIC is slightly larger (6.8 points) for model 4, meaning the more parsimonious model 2 is preferred (Raftery 1995).

We also test the contribution of segregation within the unranked programs to overall prestige segregation by constraining the scale values corresponding to the three Carnegie types to equivalence (see model 4*, Table 2). These constraints yield small but statistically significant reductions in the fit of model 4 applied to the relative prestige array ($L^2$ test statistic = 13.4, 2 df) and the absolute prestige array ($L^2$ test statistic = 15.3, 2 df). BIC, however, prefers the more parsimonious model 4* over model 4 or model 2, even for the absolute prestige array. Because the two criteria for model selection are ambiguous, we will continue to differentiate the unranked categories in later models. We note, though, that most prestige segregation occurs among ranked programs.
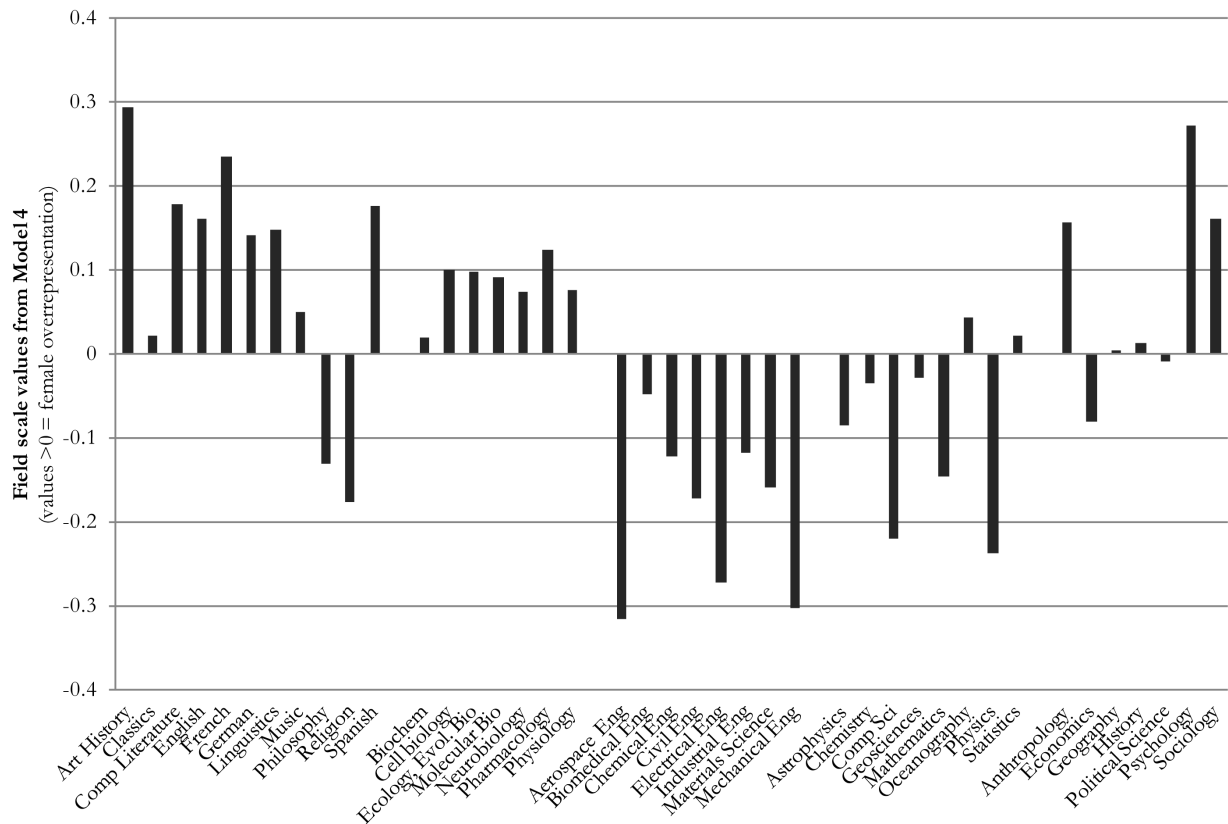
### What Is the Pattern of Field Segregation?

Figure 1 graphs the field segregation scale values from model 4 applied to the relative prestige array; scale values estimated from the absolute prestige array are nearly identical (r = 0.98). Scale values greater than zero indicate female overrepresentation relative to the average field, and scale values less than zero indicate female underrepresentation. The greater the absolute magnitude (positive or negative) of the scale value, the more segregated that field is relative to the average. These scale values show the same pattern that has been found in prior research (e.g., England et al. 2007), so we will not discuss them in depth here.

We assess the association between field segregation and skills by estimating models that fit prestige segregation, the two-way association between fields and prestige, and the two-way associations between gender and math skill groups (model 5a) and gender and verbal skill groups (model 5b). Applied to the relative prestige arrays, these models show that 64.6 percent of field segregation is associated with math skills and only 18.5 percent with verbal skills (see model contrasts 6 and 7, Table 2); these values are much the same in the absolute prestige array. This strong association between segregation and math skills is greater than the estimated association between segregation and the humanities–STEM divide in undergraduate education found in other research (Barone 2011), although differences in measures and methods can't be ruled out.[10]

### What Is the Pattern of Prestige Segregation?

Figure 2 graphs the prestige segregation scale values from model 4, fit to the relative prestige array. The scale values, which have the same interpretation as those in Figure 1, show that men are overrepresented in programs in the top three deciles but especially in the top decile. Women's representation increases toward the middle of the prestige distribution, reaching its peak (among ranked programs) in the 71–80th percentile bucket.[11] Men's representation increases in the bottom two prestige groups, such that these two deciles show male overrepresentation (i.e., scale values below 0), although not to the same extent as male overrepresentation in the top two decile groups. This modestly curvilinear pattern breaks down in unranked programs, where women are overrepresented, in particular in the lower research or nonresearch universities ("DRU/S/O"). The latter is, we suspect, partly a function of the large number of programs and graduates in psychology in the DRU/S/O,
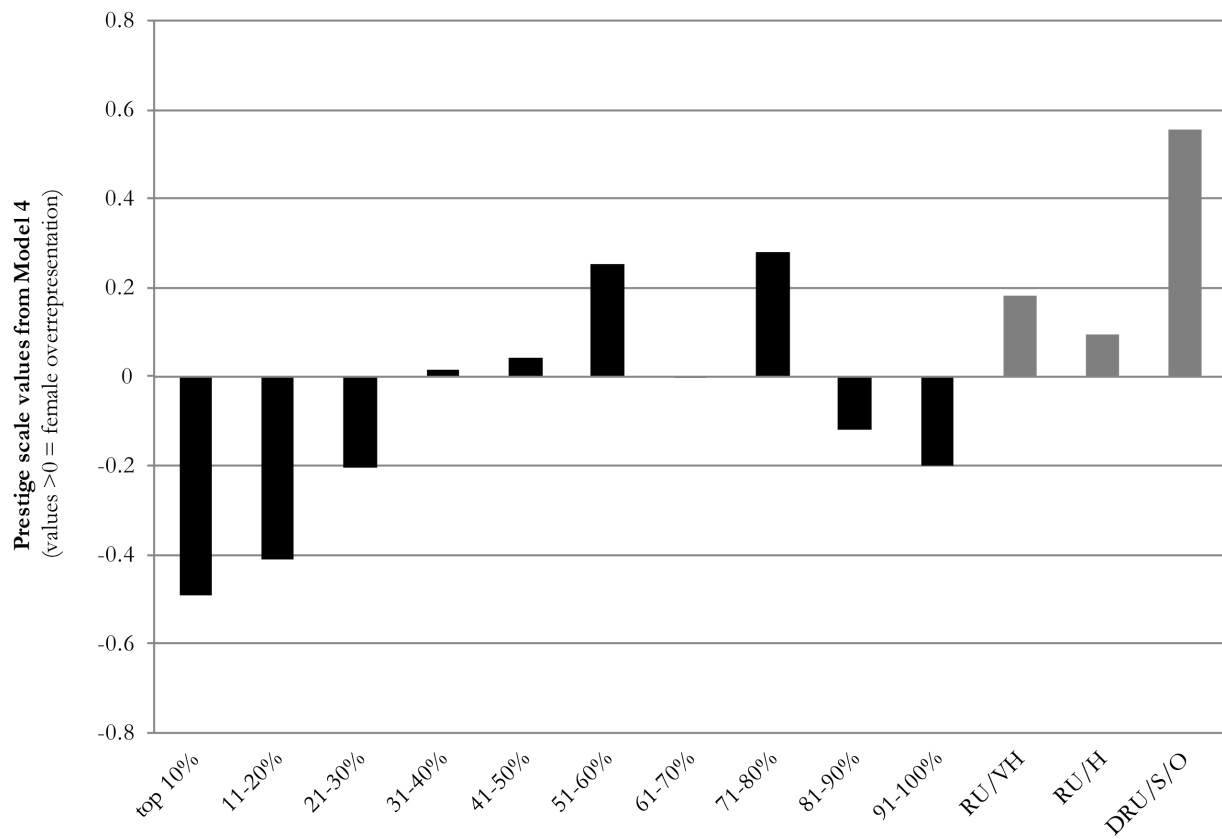
**Figure 1:** Pattern of field segregation among U.S. doctorates, 2003–2014.

Notes: Data source: IPEDs 2003–2014.

many of which are likely in clinical practice rather than research. Unfortunately, we cannot differentiate clinical and research psychology PhDs using the IPEDs data, but weighting Psychology out of the analyses eliminates the spike in female overrepresentation in the DRU/S/O category.

A different pattern emerges when we apply model 4 to the absolute prestige array (see Figure 3). As in the relative prestige arrays, men are overrepresented in the top prestige groups, although this overrepresentation is slightly lower in the top five programs than in the sixth through 10th ranked and 11th through 15th ranked programs. Women's representation increases as ranking declines, and in the lowest-ranked group (i.e., programs ranked 31st or higher), women are slightly overrepresented. Consequently, the absolute prestige array shows no evidence of the curvilinear pattern revealed in the relative prestige array (compare Figures 2 and 3). This difference emerges because the lowest absolute prestige bucket is a heterogeneous mix of middle-status programs (in fields with many ranked programs) and low-status programs (in fields with few ranked programs).
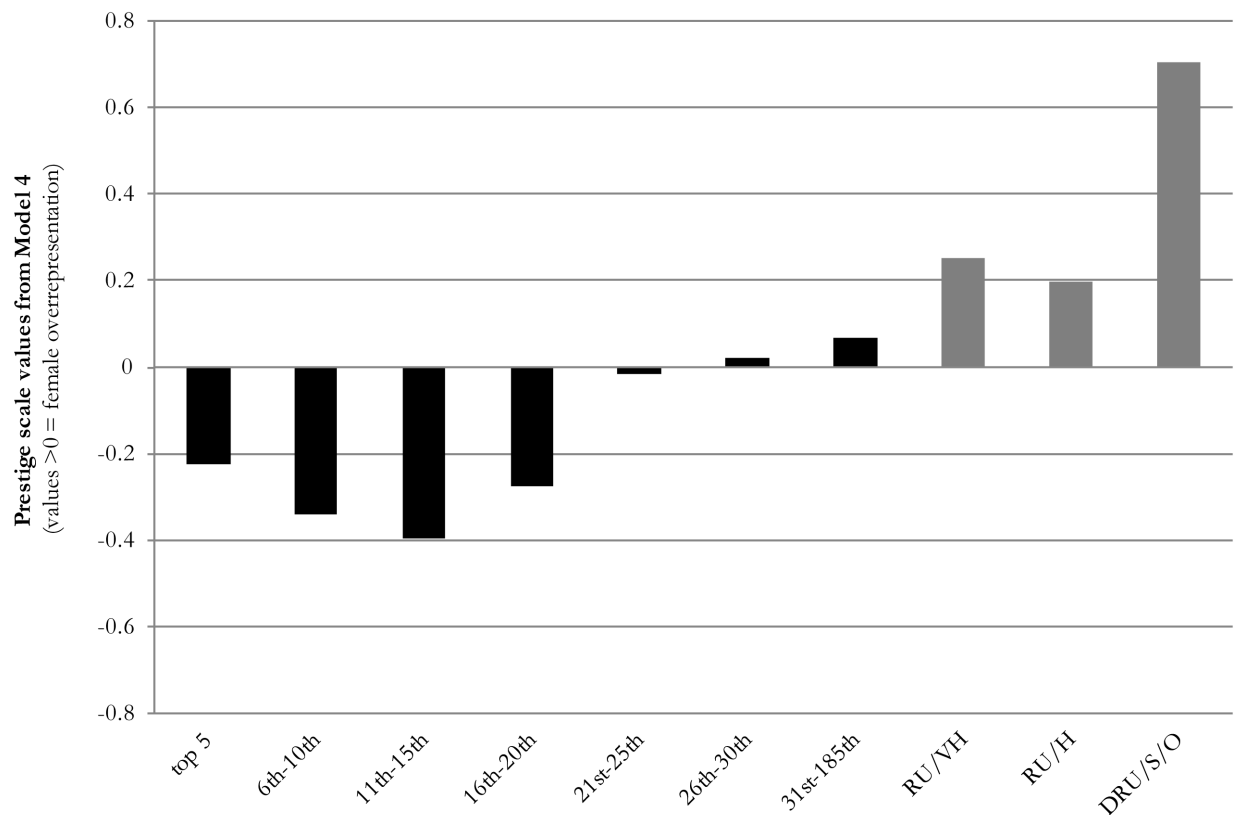
**Figure 2:** Pattern of relative prestige segregation among U.S. doctorates, 2003–2014.

Notes: Lighter shaded bars are unranked programs. Data source: IPEDs 2003–2014.

### How Does Prestige Segregation Vary across Fields?

Model 4 may, of course, mask cross-field variation in both the strength and the pattern of prestige segregation. To estimate variation in the strength of prestige segregation across fields, we fit a shift effect model (model 6; see Xie 1992; Weeden and Sørensen 2004, model 8.6) that assumes generic patterns of field and prestige segregation but allows the strength of prestige segregation to vary across fields. Model 6 captures cross-field variability with 41 field-specific "shift effects," which stretch out (in fields with stronger prestige segregation) or contract (in fields with weaker segregation) the common pattern. It improves the fit of model 4, reducing the $L^2$ by 1,197 (40 df) in the relative prestige array and 1,057 $L^2$ (40 df) in the absolute prestige array (see model contrast 9, Table 3); BIC also declines by 680.4 and 530.2 points, and indeed by the BIC criterion, model 6 is preferred over all others. Substantively, the model contrasts imply that just less than half (1,197 / 2,558 = 0.468 and 1,057 / 2,136 = 0.490) of the residual association in model 4 is attributable to cross-field differences in the strength of prestige segregation, assuming a common pattern. We will discuss specific field-level differences in segregation in the context of our final model, below.

**Figure 3:** Pattern of absolute prestige segregation among U.S. doctorates, 2003–2014.

Notes: Lighter shaded bars are unranked programs. Data source: IPEDs 2003–2014.

To assess how much of this cross-field variation is associated with math and verbal skills, we estimate variants of model 6 that replace the 41 field shift effects with five shift effects, one for each math skill group (model 7a) or verbal skill group (model 7b). The fit statistics of these models show that little cross-field variation in the strength of prestige segregation maps is attributable to skills. For example, in the relative prestige array, model 7a (math skills) yields an $L^2$ reduction of 181.9 (4 df) from model 4 (see Table 3, model contrast 6), compared to the $L^2$ reduction of 1,197 (40 df) when all 41 field-specific shift effects are fit. Put differently, math skills account for about 15 percent (182 / 1,197 = 0.152) of the total field-level variation in the strength of prestige segregation, while verbal skill shift effects account for about 22 percent (265 / 1,197 = 0.221). Skill shift effects capture slightly more cross-field variation in the absolute prestige models, but never exceed 27 percent. Field-level skills thus contribute only modestly to observed differences in the strength of prestige segregation.

**Table 3:** Fit statistics for multiplicative shift effect models of prestige segregation.

| | | Relative Prestige (10 Decile Groups + 3 Unranked) | | | | Absolute Prestige (7 Prestige Buckets + 3 Unranked) | | | |
|---|---|---|---|---|---|---|---|---|---|
| A. | Models | L2 | d.f. | Δ | BIC | L2 | d.f. | Δ | BIC |
| 4 | Field and prestige segregation | 2,559 | 480 | 2.5 | –3,641 | 2,136 | 360 | 2.1 | –2,514 |
| 6 | Model 4 + field shift effects | 1,362 | 440 | 1.9 | –4,321 | 1,089 | 320 | 1.5 | –3,044 |
| 7a | Model 4 + math skill shift effects | 2,377 | 476 | 2.4 | –3,771 | 1,915 | 356 | 2.0 | –2,683 |
| 7b | Model 4 + verbal shift effects | 2,294 | 476 | 2.4 | –3,854 | 1,849 | 356 | 2.0 | –2,749 |

| | | Change in L2 | Change in d.f. | Explained variability (%) | Change in BIC | Change in L2 | Change in d.f. | Explained variability (%) | Change in BIC |
|---|---|---|---|---|---|---|---|---|---|
| B. | Model Contrasts | | | | | | | | |
| 8. | Field shift effects on prestige segregation (model 6 vs. model 4) | 1,197 | 40 | 46.8 | –680 | 1,047 | 40 | 49.0 | –530 |
| 9 | Math skill shift effects on prestige segregation (model 7a vs. model 4) | 182 | 4 | 15.2 | –130 | 221 | 4 | 21.1 | –169 |
| 10 | Verbal skill shift effects on prestige segregation (model 7b vs. model 4) | 265 | 4 | 22.2 | –214 | 287 | 4 | 27.4 | –235 |

Δ measures the percentage of cases misclassified under the relevant model.
NOTES: Data are from the IPEDS, 2003–2014. Relative prestige N = 406,721; Absolute prestige N = 406,726.

### How Do Fields Vary in Their Pattern of Segregation?

The residual association in model 6 implies that slightly more than half of the cross-field variation in prestige segregation occurs in the underlying pattern of segregation. To explore these cross-field variations, we fit a saturated model that we parameterize to pull out cross-field variations in the strength of prestige segregation. We fit this model to the relative prestige array, given it is better at capturing heterogeneity among the lower-prestige programs. The saturated model generates 574 estimated parameters, which we present in online supplement Table S3 and graph for six illustrative fields in the humanities, social sciences, and sciences in Figure 4. The "phi" values indicate the overall strength of segregation, and the graphed scale values indicate the pattern. So, for example, English (Figure 4a) has comparatively weak segregation (phi = 0.23) but a fairly linear pattern of segregation, with male overrepresentation in top programs and female representation increasing in lower-ranked programs, whereas Sociology (Figure 4d) has a moderate level of segregation (phi = 0.47) but near gender parity in the top-ranked programs, male overrepresentation in lower-ranked programs, and strong female overrepresentation in unranked programs.

To facilitate interpretation across all 41 fields, we classify each field-by-prestige cell as male-dominated if men are overrepresented by a factor of 1.05 or greater,
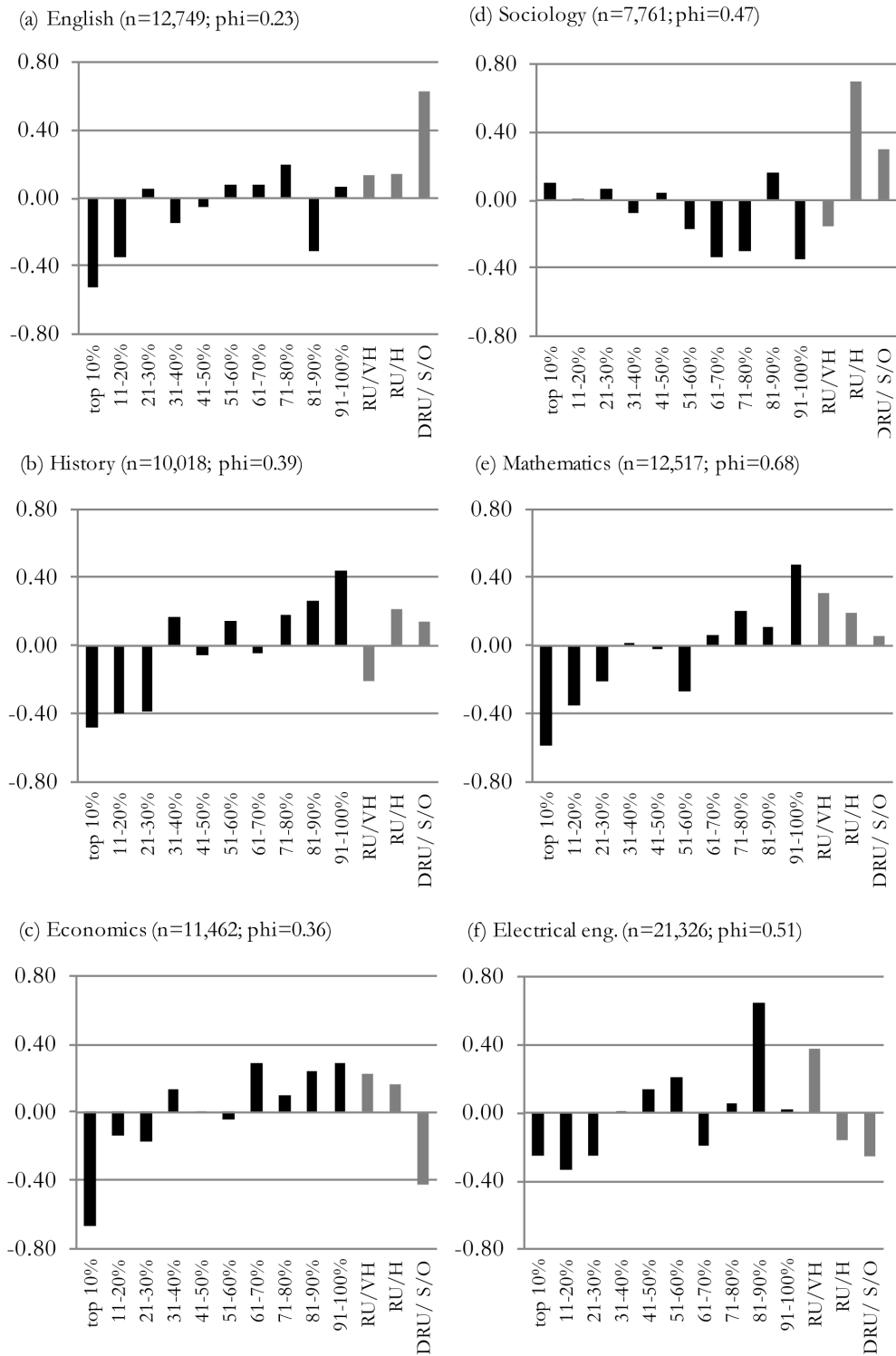
**Figure 4:** Prestige segregation scale values from saturated model, selected fields.

Notes: Phi indicates the overall strength of segregation, and positive values indicate female overrepresentation. Lighter shaded bars are unranked programs. Data source: IPEDs 2003–2014.

female-dominated if women are overrepresented by a factor of 1.05 or greater, and neutral otherwise (Weeden and Sørensen 2004). The saturated model shows that women are underrepresented in the highest prestige programs in most, but not all, fields. More specifically, top decile programs are male-dominated in 27 of the 41 NRC-ranked fields, female-dominated in four fields (Spanish, Biomedical Engineering, Materials Engineering, and Geography), and gender-neutral in the remaining 10 fields, according to our rule-of-thumb classification. However, this tally masks differences across fields in the extent of male overrepresentation in the highest prestige programs. For example, male overrepresentation in the top decile programs is quite strong in Economics (Figure 4c), where men are overrepresented by a factor of 1.27, and Mathematics (Figure 4e), where men are overrepresented by a factor of 1.48; by contrast, female overrepresentation in the top decile programs ranges from 1.08 (Spanish) to 1.16 (Biomedical Engineering). And, among the four fields in which top decile programs are female-dominated, only in Biomedical Engineering are women also overrepresented in the second decile program.

Women's underrepresentation in low-prestige programs is much less robust across fields. Using the same rule-of-thumb characterization of field-specific program cells, we find that 14 low-prestige programs are male-dominated, eight are gender-neutral, and the remaining 19 are female-dominated. Put differently, the curvilinear pattern of prestige segregation observed in Figure 2 is far from universal across fields. Finally, the saturated model illustrates the comparatively weak relationship between field skills and prestige segregation. Figure 4 shows, for example, that although women are overrepresented in the bottom two deciles in Mathematics (Figure 4d) and Economics (Figure 4e), two math-intensive fields, they are also underrepresented in the bottom two deciles in History (Figure 4c), a verbal-intensive field.

## Discussion

This article provides a systematic, multidimensional analysis of field and prestige segregation by gender in doctoral education using a unique matched data set that we constructed from the IPEDs, NRC, and GRE. We show that field segregation in doctoral education is pronounced, follows a similar pattern as field segregation at the baccalaureate level, and is strongly associated with field-level skills. Indeed, close to two-thirds of the net association between gender and field is captured by a five-category measure of math skills. By contrast, Barone (2011) found that less than half of gender segregation by undergraduate field of study is attributable to the humanities–STEM divide. This disparity could, of course, reflect differences in the student populations (primarily undergraduate vs. exclusively graduate), in the measures of skill, or in modeling strategies. If there is indeed more skill-based gender segregation in graduate education than in undergraduate education, the sources of these disparities warrant further research.

Our core result, though, is that prestige segregation is weaker than field segregation but substantively important. On average, between 11 and 13 percent of female doctoral students would need to "trade" programs with men in order to eliminate prestige segregation (Table 1). Averaged across all fields and adjusting for field

segregation, men are overrepresented in the highest-prestige programs by a factor of 1.06. In many fields, however, male overrepresentation in the highest-prestige programs is substantially higher and in the most segregated field (Mathematics) approaches 1.5 (see Figures 2 and 3; online supplement Table S3). A six percent male advantage in elite representation in the average program, up to a 50 percent advantage in some especially prestige-segregated fields, is a nontrivial gender disparity.

We also found some evidence, albeit less robust, of a curvilinear pattern of segregation: averaged across fields, men are overrepresented in low-prestige programs as well as in high-prestige programs. This pattern does not characterize all fields, and it breaks down in the unranked programs, where women are strongly overrepresented. Moreover, although there is significant cross-field heterogeneity in both the strength and the pattern of prestige segregation, prestige segregation does not map onto field-level skills. For example, math-intensive disciplines do not necessarily have more prestige segregation than verbal-intensive disciplines, nor are they more likely to evince the curvilinear pattern shown in Figure 2.

How can these patterns of prestige segregation be understood with respect to the potential sources of prestige segregation? Like most quantitative studies of gender segregation, our data are at the aggregate level and can't be used to test mechanisms. We can, however, comment on whether the aggregate patterns we observe are consistent with aggregate patterns implied by the five social processes: self-selection based on observed ability, self-selection based on self-assessed ability, self-selection based on prestige-linked program attributes, prestige-linked admissions decisions by the programs themselves, and gender-specific attrition.

The overrepresentation of men in the highest prestige programs is broadly consistent with all of the posited mechanisms at the point of admissions. However, we did not find evidence that prestige segregation is higher in math fields, as is anticipated by the "measured ability" argument (given larger gender gaps at the right tail of the distribution on observed measures of math ability than of verbal ability). Men's overrepresentation in the top programs—and the absence of strong skill-based variation across fields in this pattern—is more consistent with self-selection based on perceived ability, at least under the assumption that there is a generalized cultural belief that men are better at all higher cognitive tasks, not just math-related tasks. However, it's also consistent self-selection based on prestige-linked program attributes or prestige-linked admissions decisions.

The curvilinear pattern of segregation is more of a puzzle. At first blush, it's tempting to attribute this pattern to greater male variance in ability. This explanation falls short, though, if the pool of applicants to graduate school is drawn from the right side of the distribution on easily observed measures of ability and achievement and on the unobserved measures with which they are correlated. Moreover, the variance of GRE scores within fields is not greater for men than it is for women (ETS 2001, 2008). Similarly, the curvilinear pattern is not consistent with selection based on self-assessed ability, which predicts the clumping of women of high ability in lower-ranked programs. It is more consistent with MSC, which argues that low-status organizations have little to lose by failing to conform, and middle-status programs are most likely to conform to institutional pressures to diversify. However,

MSC does not predict women's overrepresentation in unranked programs, nor does it anticipate the substantial variation in the curvilinear pattern across fields. The curvilinear pattern of prestige segregation certainly bears further investigation, not only to understand why it emerges but also to understand why it is stronger in some fields than others.

Regardless of its source, the basic pattern of prestige segregation will be familiar to students of gender inequality: women are underrepresented among graduates of programs that most often lead to the higher paying, higher prestige jobs. This pattern has obvious implications for efforts to address gender inequality in the STEM workforce, including academia. Indeed, representatives of elite STEM departments have long claimed that a barrier to diversifying the faculty is the shortage of women (and minority) PhDs from status-equivalent institutions (e.g., Hopkins 2006). Our results show that in most fields, the tacit assumption—that elite PhD pipelines are more male-dominated than average PhD pipelines—is on the mark. From a policy perspective, this implies that efforts to diversify the faculty at elite research institutions must be complemented by efforts to reduce prestige segregation at the doctoral level.

Our analysis also holds two lessons for research on gender segregation in higher education. First, the near exclusive focus in the theoretical literature on the social psychological, macroinstitutional, and cultural antecedents of segregation might usefully be supplemented with attention to the organizational antecedents of segregation. In particular, degree-granting programs and universities are organizational actors that operate within local status orders and within institutional rules that encourage status-seeking admissions practices (Sauder and Lancaster 2006; Espeland and Sauder 2007; Stevens 2007). Second, as important as fields of study are for understanding the experiences of men and women in higher education, they do not capture all the ways that men and women's experiences in higher education, even within a given education level, differ. There are many other structural positions in higher education—including but not limited to program prestige—across which men and women may be segregated, potentially with important consequences for gender inequality in higher education.

## Notes

1   We exclude for-profit graduate programs, professional degrees, business degrees, and education degrees from our analysis. The processes that lead to segregation in these programs likely differ from those that lead to segregation in research fields.

2   We don't make claims about the source of gender differences in GRE test scores. We note, however, that these gender differences are probably affected by self-selection into the test, which may in turn be affected by gender-differentiated perceptions of the payoff to a PhD and labor market alternatives.

3   If women hold themselves to stricter standards than men, in theory women could self-select out of the graduate school applicant pool altogether. We see little evidence of this in aggregate data on GRE takers: in 2001 and 2002, nearly 60 percent of GRE test takers were women, which exceeds their share of doctoral recipients in the late 2000s.

4  Given suitable data, it would be possible to estimate how much of the gender–prestige association is captured by funding, faculty gender composition, and other program attributes. The IPEDS contains some of these measures for universities but not for individual programs within universities.

5  We use "status" in this section to be consistent with the MSC literature but assume status and prestige are interchangeable.

6  Field segregation changed very little over this period (NSF 2015b:4; see also England et. al 2007). We assume the same is true of prestige segregation.

7  The NRC ranks six PhD programs at the Uniformed Services University of Health Sciences, which does not report to IPEDS.

8  Discretizing the GRE scores allows for nonlinearities in the relationship between each type of skill and prestige segregation and simplifies the presentation of the models.

9  Because we will discuss model 4's estimates at length, we give its formula as follows: $m_{ijk} = \alpha \beta_i \gamma_j \delta_k \lambda_{ik} e^{\phi(Z_j \mu_i)} e^{\psi(Z_j \nu_k)}$, where $i$ indexes field, $j$ indexes sex, $k$ indexes prestige group, $\phi$ captures the strength of field segregation, $\mu i$ are the scale values that define the pattern of field segregation, $\psi$ captures the strength of prestige segregation, and the $\nu k$ are prestige scale values that the pattern of prestige segregation. Scale values are identified under the standard normalization constraint of zero mean and unit variance.

10  Barone (2011) uses a rule-of-thumb, binary indicator of whether a field is humanities or STEM.

11  The dip in the scale value for the sixth decile is present in supplementary analyses we conducted to check robustness: weighting out Psychology, excluding the 284 unmatched programs, and imposing equality constraints on segregation in the three unranked programs. Estimates from the saturated model (see below; also online supplement Table S3) show that in most fields, the sixth decile is female-dominated or gender-neutral, but in a handful of fields it is heavily male-dominated.

# References

Ampaw, Frim D., and Audrey J. Jaeger. 2012. "Completing the Three Stages of Doctoral Education: An Event-History Analysis." *Research in Higher Education* 53: 640-660. https://doi.org/10.1007/s11162-011-9250-3

Baker, Joe G. 1997. "Gender, Race and PhD Completion in Natural Science and Engineering." *Economics of Education Review.* 17(2): 179-188. https://doi.org/10.1016/S0272-7757(97)00014-9

Barone, Carlo. 2011. "Some Things Never Change: Gender Segregation in Higher Education Across Eight Nations and Three Decades." *Sociology of Education* 84(2): 157-176. https://doi.org/10.1177/0038040711402099

Berg, Helen M., and Marianne A. Ferber. 1983. "Men and Women Graduate Students: Who Succeeds and Why?" *Journal of Higher Education* 54(6):629-648. https://doi.org/10.2307/1981934

Bobbitt-Zeher, Donna. 2007. "The Gender Income Gap and the Role of Education." *Sociology of Education* 80(1): 1-22. https://doi.org/10.1177/003804070708000101

Burris, Val. 2004. "The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks." *American Sociological Review* 69(2):239-264. https://doi.org/10.1177/000312240406900205

Charles, Maria, and Karen Bradley. 2002. "Equal but Separate? A Cross-national Study of Sex Segregation in Higher Education," *American Sociological Review*, 67(4):573-99. https://doi.org/10.2307/3088946

Charles, Maria, and Karen Bradley. 2009. "Indulging our Gendered Selves: Sex Segregation by Field of Study in 44 Countries." *American Journal of Sociology,* 114(4):924-976. https://doi.org/10.1086/595942

Charles, Maria, and David B. Grusky. 1995. "Models for Describing the Underlying Structure of Sex Segregation." *American Journal of Sociology* 100:931-71. https://doi.org/10.1086/230605

Charles, Maria, and David B. Grusky. 2004. *Occupational Ghettos: The Worldwide Segregation of Men and Women.* Palo Alto, CA: Stanford University Press.

Clance, Pauline R., and Maureen Ann O'Toole. 1988. "Impostor Phenomenon: An Internal Barrier to Empowerment and Achievement." *Women and Therapy* 6:51-64. https://doi.org/10.1300/J015V06N03_05

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106:1691-1730. https://doi.org/10.1086/321299

Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69:93-113. https://doi.org/10.1177/000312240406900106

Davies, Scott, and Neil Guppy. 1997. "Fields of Study, College Selectivity, and Student Inequalities in Higher Education." *Social Forces* 75(4):1417-1438. https://doi.org/10.1093/sf/75.4.1417

Dwyer, Rachel, Randy Hodson, and Laura McCloud. 2013. "Gender, Debt, and Dropping out of College." *Gender and Society* 27(1):30-55. https://doi.org/10.1177/0891243212464906

Educational Testing Service. 2001. *Sex, Race, Ethnicity, and Performance on the GRE ® General Test, 2001-2002.* ETS: Iowa City, Iowa.

Educational Testing Service. 2008. *Guide to the Use of Scores, 2008-2009.* ETS: Iowa City, Iowa.

Ehrenberg, Ronald G., and Panagiotis G. Mavros. 1995. "Do Doctoral Students' Financial Support Patterns Affect their Times-to-Degree and Completion Probabilities?" *Journal of Human Resources* 30:581-609. https://doi.org/10.2307/146036

England, Paula, and Su Li. 2006. "Desegregation Stalled: The Changing Gender Composition of College Majors, 1971-2002." *Gender & Society* 20(5): 657-77. https://doi.org/10.1177/0891243206290753

England, Paula, Paul Allison, Su Li, Noah Mark, Jennifer Thompson, Michelle Budig, and Han Sun. 2007. "Why are Some Academic Fields Tipping Toward Female? The Sex Composition of US Fields of Doctoral Degree Receipt, 1971-2002. *Sociology of Education* 80:23-42. https://doi.org/10.1177/003804070708000102

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113:1-40. https://doi.org/10.1086/517897

Foschi, Martha. 2008. "Gender, Performance Level, and Competence Standards in Task Groups." *Social Science Research* 38:447-57. https://doi.org/10.1016/j.ssresearch.2008.10.004

Fox, Mary Frank. 1995. "Women and Scientific Careers." Pp. 205-223 in *Handbook of Science and Technology Studies,* edited by S. Jasanoff, J. Markle, J. Petersen, and T. Pinch. Thousand Oaks: Sage. https://doi.org/10.4135/9781412990127.n10

Gilford, Dorothy M., and Joan Snyder. 1977. *Women and Minority PhDs in the 1970's: A Data Book.* Washington, DC: National Research Council/National Academy of Sciences.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118(3):1049-74. https://doi.org/10.1162/00335530360698496

Hale, Galina, and Tali Regev. 2014. "Gender Ratios at Top PhD Programs in Economics." *Economics of Education Review* 41(August):55-70. https://doi.org/10.1016/j.econedurev.2014.03.007

Hopkins, Nancy. 2006. "Diversification of a University Faculty: Observations on Hiring Women Faculty in the Schools of Science and Engineering at MIT." MIT Faculty Newsletter XVIII(4). Last accessed Nov. 2016, http://web.mit.edu/fnl/volume/184/hopkins.html.

Jacobs, Jerry A. 1995. "Gender and Academic Specialties." *Sociology of Education* 68: 81-98. https://doi.org/10.2307/2112776

Jacobs, Jerry A. 1999. "Gender and the Stratification of Colleges." *Journal of Higher Education*, 70(2): 161-87. https://doi.org/10.2307/2649126

Long, J. Scott. 2001. *From Scarcity to Visibility: Gender Differences in the Careers of Doctoral Scientists and Engineers.* Washington, DC: National Research Council/National Academies Press

Long, J. Scott, and Mary Frank Fox. 1995. "Scientific Careers: Universalism and Particularism." *Annual Review of Sociology*, Vol. 21 pp.45-71. https://doi.org/10.1146/annurev.so.21.080195.000401

Lott, Joe L. II, Susan Gardner, and Daniel A, Powers. 2009. "Doctoral Student Attrition in the STEM Fields: An Exploratory Event History Analysis." *Journal of College Student Retention* 11(2): 247-66. https://doi.org/10.2190/CS.11.2.e

Makel, Matthew C., Jonathan Wai, Kristen Peairs, and Martha Putallaz. 2016. "Sex Differences in the Right Tail of Cognitive Abilities: An Update and Cross-Cultural Extension." *Intelligence* 59 (November-December): 8-15. https://doi.org/10.1016/j.intell.2016.09.003

Mann, Allison, and Thomas A. DiPrete. 2013. "Trends in Gender Segregation in the Choice of Science and Engineering Majors." *Social Science Research* 42(6): 1519-41. https://doi.org/10.1016/j.ssresearch.2013.07.002

Mann, Allison, and Thomas A. DiPrete. 2016. "The Consequences of National Math and Science Performance Environment for Gender Differences in STEM Aspiration." *Sociological Science* July 12, 2016. https://doi.org/10.15195/v3.a25

Milkman, Katherine L., Modupe Akinola, and Dolly Chugh. 2012. "Temporal Distance and Discrimination: An Audit Study in Academia." *Psychological Science* 23(7): 710-717. https://doi.org/10.1177/0956797611434539

Moss-Racusin, Corrine A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences* 109(41): 16474–16479. https://doi.org/10.1073/pnas.1211286109

National Research Council. 2010. *A Data-Based Assessment of Research-Doctorate Programs in the United States.* The National Academies Press.

National Science Foundation. 2015a. *Doctoral Recipients from US Universities: 2014.* Data Tables, Table 14. http://www.nsf.gov/statistics/2016/nsf16300/data/tab14.pdf, last accessed July 2, 2016.

National Science Foundation. 2015b. *"Women, Minorities, and Persons with Disabilities in Science and Engineering: 2015."* http://www.nsf.gov/statistics/2015/nsf15311/digest/nsf15311-digest.pdf. Last accessed September 29, 2016.

Nettles, Michael T., and Catherine M. Millett. 2006. *Three Magic Letters: Getting to Ph.D.* Baltimore: The Johns Hopkins University Press.

Penner, Andrew M., and Marcel Paret. 2008. "Gender Differences in Mathematics Achievement: Exploring the Early Grades and the Extremes." *Social Science Research* 37: 239-53. https://doi.org/10.1016/j.ssresearch.2007.06.012

Phillips, Damon J., and Ezra W. Zuckerman. 2001. "Middle-Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets." *American Journal of Sociology* 107(2): 379-429. https://doi.org/10.1086/324072

Posselt, Julie R. 2016. *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping.* Cambridge: Harvard University Press. https://doi.org/10.4159/9780674915640

Raftery, Adrian T. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111-163. https://doi.org/10.2307/271063

Ransom, Michael R. 1990. "Gender Segregation by Field in Higher Education." *Research in Higher Education* 31(5):477-94. https://doi.org/10.1007/BF00992714

Riegle-Crumb, Catherine, Barbara King, Eric Grodsky, and Chandra Muller. 2012. "The More Things Change, the More They Stay the Same? Prior Achievement Fails to Explain Gender Inequality in Entry into STEM College Majors Over Time." *American Educational Research Journal* 49(6):1048-1073. https://doi.org/10.3102/0002831211435229

Ridgeway, Cecilia L. 1997. "Interaction and the Conservation of Gender Inequality: Considering Employment." *American Sociological Review* 62: 218-235. https://doi.org/10.2307/2657301

Rosser, Sue V. 2004. *The Science Glass Ceiling: Academic Women Scientists and the Struggle to Succeed.* New York: Routledge. https://doi.org/10.4324/9780203354612

Sauder, Michael, and Ryon Lancaster. 2006. "Do Rankings Matter? The Effects of *US News and World Report* Rankings on the Admissions Process of Law Schools." *Law and Society Review* 40(1): 105-134. https://doi.org/10.1111/j.1540-5893.2006.00261.x

Schiebinger, Londa, Andrea Davies Henderson, and Shannon K. Gilmartin. 2008. *Dual-Career Academic Couples: What Universities Need to Know.* Stanford: Michelle R. Clayman Institute for Gender Research.

Stevens, Mitchell. 2007. *Creating a Class: College Admissions and the Education of the Elites.* Cambridge: Harvard University Press.

Wagner, David G. and Joseph Berger. 1997. "Gender and Interpersonal Task Behaviors: Status Expectation Accounts." *Sociological Perspectives* 40(1): 1-32. https://doi.org/10.2307/1389491

Wallace, Jean E., and Valerie A. Haines. 2004. "The Benefits of Mentoring for Engineering Students." *Journal of Women and Minorities in Science and Engineering* 10(4): 377-91. https://doi.org/10.1615/JWomenMinorScienEng.v10.i4.60

Weeden, Kim A., and Jesper B. Sørensen. 2004. "A Framework for Analyzing Industrial and Occupational Segregation." Pp. 245-96 in *Occupational Ghettos,* by Maria Charles and David Grusky. Stanford: Stanford University Press.

Xie, Yu. 1992. "The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables." *American Sociological Review* 57:380-395. https://doi.org/10.2307/2096242

Xie, Yu, and Kimberlee A. Shauman. 2003. *Women in Science.* Cambridge, MA: Harvard University Press.

Zwick, Rebecca. 1991. *An Analysis of Graduate School Careers in Three Universities.* Princeton: Educational Testing Service.

**Kim A. Weeden:** Department of Sociology, Cornell University.
E-mail: kw74@cornell.edu.

**Sarah Thébaud:** Department of Sociology, University of California - Santa Barbara.
E-mail: sthebaud@gmail.com.

**Dafna Gelbgiser:** Center for the Study of Inequality, Cornell University.
E-mail: dg432@cornell.edu.