Leaks in the pipeline: separating demographic inertia from ongoing gender differences in academia

Electronic Supplementary Material

Allison K. Shaw[1*†] and Daniel E. Stanton[1,2*†]

[1] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

[2] Division of Plant Sciences, Research School of Biology, the Australian National University, Acton ACT 0200, Australia

*Authors for correspondence (akshaw@princeton.edu; daniel.stanton@anu.edu.au).

†These authors contributed equally to this work.

Electronic Supplementary Material: Detailed Methods

All simulations and calculations were done using *Matlab*. M-files are available upon request from the authors.

Creation of pipeline

We modeled the academic career pipeline as a series of 5 pools (undergraduate studies, graduate studies, postdoctoral fellowships, assistant professorship and tenured professorship) where scholars are transferred between pools at a rate inversely proportional to the average length of time spent in the pool (equivalent to a turnover time). Individuals moving out of one pool were eligible to move up and fill empty positions in the next pool, and those that did not move up left the pipeline.

We generated the structure of the pipeline model from National Science Foundation data as follows. Since the number of academics in the U.S. has dramatically increased over the past thirty years, we set the model pool sizes as the actual number of graduate students, postdoctoral fellows, assistant professors, and tenured professors (annual National Science Foundation data from 1979-2006; NSB 10-01, NSF 10-307). We did not model an undergraduate pool explicitly, but rather used the number of bachelor's degrees awarded in a given year (NSF 08-321) to calculate the number of potential graduate student candidates, as well as to identify a loss rate associated with the transition to graduate school. Due to the absence of data on postdoctoral fellows prior to 1993, we estimated the number of fellows between 1979 and 1992 based on a linear regression of the number of fellows from 1993 to 2006. Due to the rapid turnover rate of this pool, any offset caused by this assumption was overcome within the first few years of the simulation, and therefore is unlikely to have had any lasting effects on the model results.

We set the proportion of undergraduate degrees awarded to women for each year of the simulation to National Science Foundation data values. For the other pools (graduate students, postdoctoral fellows, assistant professors, and tenured professors), we set initial female participation as the National Science Foundation data values from 1979 (NSF 08-308, NSF 09-305, NSF 10-307), and simulated female participation for subsequent years. We started simulations in the year 1979 since National Science Foundation data on pool sizes were not available for earlier years. In longer career stages, the use of a single value for female participation is unrealistic, as retirement or promotion is primarily from the older individuals. We therefore partitioned the graduate student and tenured professor pools into 2 and 5 sub-pools respectively.

We constructed a pipeline model for all Science and Engineering (S & E), as well as for several disciplines: Agricultural and Biological Sciences (BIO), Engineering (ENG), Mathematics (MAT; includes mathematics and statistics), Physical Sciences (PHY; includes physics and chemistry), Psychology (PSY), and Social Sciences (SOC). Earth, Atmospheric and Ocean Sciences could not be included due to the lack of data on faculty composition in the National Science Foundation databases.

Model simulation

For each year we calculated the number of individuals moving in and out of each pool as estimated based on transition rates and changing pool sizes (figure 1). We first removed the estimated number of retiring tenured professors from that pool, filling empty slots with available assistant professors. We then filled the empty slots created in the assistant professor pool with available postdoctoral fellows, in turn filling any postdoctoral open slots with available graduating doctoral students, and lastly filled open slots in the doctoral student pool with graduating bachelors students. The actual transition rates were calculated as:

$$\mu_G(t) = N_G(t+1) - N_G(t) + \left(\frac{1}{\tau_G}\right) N_G(t) \tag{1}$$

$$\mu_P(t) = N_P(t+1) - N_P(t) + \left(\frac{1}{\tau_P}\right) N_P(t) \tag{2}$$

$$\mu_A(t) = N_A(t+1) - N_A(t) + \left(\frac{1}{\tau_A}\right) N_A(t) \tag{3}$$

$$\mu_T(t) = N_T(t+1) - N_T(t) + \rho_T(t) \tag{4}$$

$$\rho_T(t) = \left(\frac{1}{\tau_T}\right) N_T(t) \tag{5}$$

$$\lambda_U(t) = D_U(t) - \mu_G(t) \tag{6}$$

$$\lambda_G(t) = D_G(t) - \mu_P(t) \tag{7}$$

$$\lambda_P(t) = \left(\frac{1}{\tau_P}\right) N_P(t) - \mu_A(t) \tag{8}$$

$$\lambda_A(t) = \left(\frac{1}{\tau_A}\right) N_A(t) - \mu_T(t) \tag{9}$$

$$\delta_G(t) = \left(\frac{1}{\tau_G}\right) N_G(t) - D_G(t) \tag{10}$$

where $\mu_i(t)$ are the number of individuals transitioning into class i (G for graduate student, P for postdoctoral fellow, A for assistant professor and T for tenured professor) from the previous class in year t, $\lambda_i(t)$ are the number of individuals in class i that do not go on to the next class in year t (and instead leak out of the pipeline), $\delta_G(t)$ is the number of graduate students dropping out in year t (before receiving degree), and $\rho_T(t)$ is the number of tenured professors retiring in year t. The remaining parameters in equations (1-10) came from National Science Foundation data: $N_i(t)$ are the number of individuals in class i in year t (NSB 10-01, NSF 10-307), $D_i(t)$ is the number of degrees of class i awarded in year t (NSF 08-321), τ_i is the average number of years spent in each class (7.2, 7.3, 7.0, 6.9, 6.4, 7.3, 8.9 and for graduate studies in S&E, BIO, ENG, MAT, PHY, PSY and SOC, respectively; and 1.9, 2.2, 1.3, 1.7, 1.9, 1.2, and 1.2 for postdoctoral fellowships in S&E, BIO, ENG, MAT, PHY, PSY and SOC, respectively; NSB 10-01, NSF 08-307). For the results presented in this paper we assumed individuals went through two postdoctoral fellowships before moving on to become an assistant professor.

For postdoctoral fellows and assistant professors we could not distinguish between individuals leaving the pipeline in the middle and those leaving at transition points (e.g. an assistant professor leaving before, or just after, receiving tenure) so we lumped together both losses in a general loss term for each class (λ_P and λ_A). Similarly we could not distinguish between tenured professors leaving early and those retiring at the end of their career, so we lumped these together in a single term, retirement (ρ_T). However, since data is available for the number of doctoral degrees each year, we were able to distinguish between graduate students leaving the pipeline during grad school (δ_G) and those receiving doctoral degrees but leaving academia (λ_G).

In some simulation years the numbers did not quite match up and we had to make the following adjustments. When the number of empty tenured professor slots exceeded the number of assistant professors available (as estimated from turnover times), then we moved as many assistant professors as necessary into tenured slots. Similarly we moved up as many postdoctoral fellows as necessary to fill empty assistant professor slots, even when it exceeded the estimated number of available postdoctoral fellows. This was only necessary during the earliest years of the model simulations (1980's), which we believe reflects the fact that during this time it was typical to spend little or no time as a postdoctoral fellow. If the number of empty assistant professor slots exceeded the total number of existing postdoctoral fellows in that discipline in that year, then we assumed that all fellows were hired as assistant professors, and the remaining professor slots were filled with fellows from outside the pool (coming from other scientific disciplines or returning to academia after having previously left), but in the same sex ratio as the pool of postdoctoral fellows. Similarly, if the number of empty postdoctoral fellow slots exceeded the number of students receiving doctoral degrees in that year, we assumed the remaining slots were filled with doctoral graduates from outside the pool (either from other scientific disciplines or graduates from previous years that had temporarily left academia).

As mentioned, the graduate student and tenured professor pools were partitioned into two and five sub-pools, respectively. Transitions between sub-pools were estimated based on turnover time (the amount of time spent in each sub-pool within a pool was approximately equal). Graduate students leaving the pipeline before receiving a degree (δ_G) were pulled from both sub-pools (half from each), but graduate students leaving the pool with a doctoral degree were assumed to come only from the second sub-pool. Tenured professors retiring (ρ_T) were pulled only from the last sub-pool.

Once the number of individuals moving in and out of each class was calculated, we subdivided this based on the proportion of female academics in each class. Initial female participation in each pool was taken from National Science Foundation 1979 data, except for the case of postdoctoral fellows where female participation was not available before 1999. In this case we assumed an initial female participation that was the mean of the values for graduate students and assistant professors. As with pool sizes, any minor offsets created by this assumption were rapidly overcome within the first few years of the simulation. We calculated the number of women in each pool as

$$F_G(t+1) = F_G(t) - D_G(t)f_G(t) - \delta_G(t)f_G(t) + \mu_G(t)f_U(t) \quad (11)$$

$$F_P(t+1) = F_P(t) - \mu_A(t)f_P(t) - \lambda_P(t)f_P(t) + \mu_P(t)f_G(t) \quad (12)$$

$$F_A(t+1) = F_A(t) - \mu_T(t)f_A(t) - \lambda_A(t)f_A(t) + \mu_A(t)f_P(t) \quad (13)$$

$$F_T(t+1) = F_T(t) - \rho_T(t)f_T(t) + \mu_T(t)f_A(t) \quad (14)$$

where $F_i(t)$ is the simulated number of women in class i at time t, $f_i(t)$ is the proportion of individuals in class i at time t who are female, and other variables are as listed above. With our pipeline model structure in place, we then simulated the flow of individuals through the pipeline. We assumed that at each transition, all types of individuals (males and females, in this case) were equally likely to continue on in the pipeline or leak out, and therefore individuals entering a given pool were drawn from the pool below in proportion to the sex ratio in that lower pool.

Calculation of Inertial Effect

The inertial effect (IE) was calculated as the ratio between the proportion of individuals in a class and within discipline that were female from the data, to the predicted proportion from the model. This gives a metric for how much female participation can be accounted for by demographic inertia alone, where an IE value equal to 1 indicates that the data exactly match the prediction under demographic inertia alone, while a value less than one indicates that women are less represented than would be expected by demographic inertia, and a value greater than one indicates that they are more represented. The IE by class (table 1) was determined by running the pipeline model, assuming no gender differences in attrition, and starting with National Science Foundation data from 1979-2006 for the proportion of females in the previous class for which there is long-term data (UG for GR, GR for PD, GR for AP, AP for TP). The IE over time (figure 3, table S1) was determined by running the pipeline model, assuming no gender differences in attrition, and starting with National Science Foundation data on the proportion of females at the undergrad level, starting in different years and recording the simulated proportion of women 5 years later.

Electronic Supplementary Material: Additional Results

Table S1: The inertial effect (IE) over time: most disciplines and transitions show improvement towards and IE value closer to 1.

IE in undergrad to grad student transition							
years	S&E	BIO	ENG	MAT	PHY	PSY	SOC
1979-1985	0.89	1.02	1.05	0.89	0.87	1.02	1.01
1986-1992	0.91	0.98	1.03	0.92	0.87	1.00	1.01
1993-1999	0.96	1.02	1.18	1.10	0.87	1.00	1.06
2000-2006	0.91	0.97	1.12	1.17	0.83	1.00	1.01
IE in grad student to postdoc transition							
years	S&E	BIO	ENG	MAT	PHY	PSY	SOC
1979-1985	NA	NA	NA	NA	NA	NA	NA
1986-1992	NA	NA	NA	NA	NA	NA	NA
1993-1999	0.80	0.82	0.82	0.58	0.64	0.73	0.86
2000-2006	0.81	0.79	0.83	0.59	0.70	0.69	0.86
IE in grad to assistant professor transition							
years	S&E	BIO	ENG	MAT	PHY	PSY	SOC
1979-1985	0.50	0.52	0.23	0.35	0.36	0.50	0.48
1986-1992	0.66	0.68	0.41	0.31	0.43	0.55	0.57
1993-1999	0.76	0.74	0.50	0.52	0.61	0.62	0.65
2000-2006	0.83	0.78	0.70	0.51	0.74	0.72	0.75
IE in assistant to tenured professor transition							
years	S&E	BIO	ENG	MAT	PHY	PSY	SOC
1979-1985	0.95	0.98	0.89	0.90	1.00	0.92	0.96
1986-1992	1.00	1.95	0.77	0.85	0.83	1.20	0.93
1993-1999	1.10	1.08	0.90	1.20	1.17	1.04	1.11
2000-2006	1.11	1.19	1.26	0.84	1.02	1.05	1.09

Electronic Supplementary Material: References

NSB Publication 10-01. 2010 National Science Board, Science and Engineering Indicators 2010.

See http://www.nsf.gov/statistics/seind10.

NSF Publication 08-307. 2008 National Science Foundation, Division of Science Resources Statistics, Postdoc Participation of Science, Engineering, and Health Doctorate Recipients.

See http://www.nsf.gov/statistics/infbrief/nsf08307.

NSF Publication 08-308. 2008 National Science Foundation, Division of Science Resources Statistics, Thirty-Three Years of Women in S&E Faculty Positions.

See http://www.nsf.gov/statistics/infbrief/nsf08308.

NSF Publication 08-321. 2008 National Science Foundation, Division of Science Resources Statistics, Science and Engineering Degrees: 19662006.

See http://www.nsf.gov/statistics/nsf08321.

NSF Publication 09-305. 2009 National Science Foundation, Division of Science Resources Statistics, Women, Minorities, and Persons with Disabilities in Science and Engineering.

See http://www.nsf.gov/statistics/wmpd.

NSF Publication 10-307. 2010 National Science Foundation, Division of Science Resources Statistics, Graduate Students and Postdoctorates in Science and Engineering: Fall

2007.

See http://www.nsf.gov/statistics/nsf10307.