ONLINE APPENDIX TO ACCOMPNAY:

Modeling Heterogeneity in the Organization of Scientific Work

Hazhir Rahmandad and Keyvan Vakili
hazhir@mit.edu kvakili@london.edu

Table of Contents

A-Heterogeneity in outcomes across branches of science

Figures in table 1 are compiled from the National Science Foundation various data sources for 2014 and 2015. They show significant variation across funding, PhD graduates, and articles per faculty (as much as 960%, 240%, and 480% respectively), even after aggregating the data over very different institutions and large branches of science. Table 1 in the main paper shows that variations would be more pronounced if we consider more narrowly defined fields in a single institution, such as electrical engineering and biology rather than engineering and life sciences. Also included in this table is the fraction of research funds going to equipment, which is rather limited, suggesting that equipment costs cannot explain the broader variations in outcomes.

Table S1- Heterogeneity in the outcomes of U.S. scientists

RESEARCH UNIVERSITIES IN THE U.S.A.	FUNDING PER FACULTY (THOUSANDS OF $ PER YEAR)	PHD GRADUATES PER FACULTY PER YEAR	ARTICLES PER FACULTY PER YEAR	EQUIPMENT FUNDS (FRACTION OF TOTAL RESEARCH COSTS)
PHYSICAL SCIENCES	1,082	0.51	3.78	0.05
MATHEMATICAL SCIENCES	129	0.35	0.79	0.01
COMPUTER SCIENCE	780	0.74	0.86	0.04
LIFE SCIENCES	1,239	0.45	2.92	0.03
PSYCHOLOGY	135	0.45	1.07	0.02
SOCIAL SCIENCES	155	0.35	0.83	0.01
ENGINEERING	992	0.83	1.12	0.05

B-Analytical solutions

The model notations are summarized below:

Parameters (assumed constant):

r_p : The productivity of staff relative to scientist in contributing to labor component of research

t_a : The fraction of faculty time spent on training and advising one staff member

t_f : The fraction of faculty time spent on bringing the funding needed to sustain one staff member

t_e : The fraction of faculty time spent on bringing the funding needed to sustain one unit of equipment

e_s : The productivity of scientist in contributing to labor component of publication production

Decisions and Outcomes:

P: Number of funded research staff

E: Equipment

T_r: Scientist's Research Execution Time

O: Research Output

T_f: Scientist time on raising funds

T_a: Scientist time advising and managing the team

O_F: Research fraction labored by funding

1-Baseline model

To solve the first optimization model (equation 3), the Karush-Kuhn-Tucker conditions are used:

Defining: $\Lambda = (e_s T_r + e_s r_p P)^{\alpha} E^{1-\alpha} + \lambda(T_r + Pt_f + Pt_a + Et_e - 1) + \mu_1 T_r + \mu_2 E + \mu_3 P$

The following equations will hold at the local extremums for the original problem:

$$\frac{\partial \Lambda}{\partial T_r} = 0; \frac{\partial \Lambda}{\partial E} = 0; \frac{\partial \Lambda}{\partial P} = 0; \quad (4)$$

$$T_r + Pt_f + Pt_a + Et_e - 1 = 0 \quad (5)$$

$$\mu_1 T_r = 0; \mu_2 E = 0; \mu_3 P = 0; \quad (6)$$

This system of seven equations and seven unknowns can be solved by using the first three equations to find μ_1, μ_2, μ_3 in terms of other unknowns, and then solving the remaining four equations to find T_r, P, and E. This procedure provides two possible answer sets that maximize the original objective function depending on the parameter settings in the model. These results are summarized in Table 2.

Table S2 – Metrics for the extent of relationship between a research group and external clients.

	Funded Research Model If $r_p > t_a + t_f$	Hands-on research Model If $r_p \leq t_a + t_f$
Number of funded PhD Students (P)	$\alpha/(t_a + t_f)$	0
Equipment (E)	$(1-\alpha)/t_e$	
Scientist's Research Execution Time (T_r)	0	α
Research Output (O)	$(e_s r_p P)^{\alpha}(\frac{1-\alpha}{t_e})^{1-\alpha}$	$(e_s \alpha)^{\alpha}(\frac{1-\alpha}{t_e})^{1-\alpha}$
Scientist time on raising funds (T_f)	$1-\alpha t_a/(t_a + t_f)$	$1-\alpha$
Scientist time advising and managing the team (T_a)	$\alpha t_a/(t_a + t_f)$	0
Research fraction labored by funding (O_F)	1	0

2- Heterogeneous research tasks

In the main analysis we focused on a single type of research task. Here we extend the analysis by separating research design and research implementation into two different tasks, where the former can only be done by the scientist. We can explicitly capture research design time (T_D) as a variable the scientist can change, and consider its impact on the research output by moderating the scientist's time allocation problem to:

$$\underset{T_r,P,E,T_D}{Maximize}:(e_sT_r+e_sr_pP)^{\alpha}{T_D}^{\beta}E^{1-\alpha-\beta} \quad (8)$$

Subject to:

$$T_r+Pt_f+Pt_a+T_D+Et_e=1$$

$$0\le P,E,T_r,T_D$$

Solving this problem, using a similar method, leads to very similar results, where a fixed fraction of scientist's time ($T_D=\beta$) will be allocated to research design, and the optimum equipment level is $E=1-\alpha-\beta$. This solution is consistent with our original assumption, so we used the simpler model in the text.

3- Institutional support and internal funding

Finally, the possibility of using internal funding, fellowships and teaching assistantships to fund research staff can be included. The resulting system of equations is thus modified to:

$$\underset{T_r,P,E,G}{Maximize}:(e_sT_r+e_sr_p(P+G(1-t_g)))^{\alpha}E^{1-\alpha} \quad (9)$$

Subject to:

$$T_r+Pt_f+Pt_a+Gt_a+Et_e=1$$

$$G\le g$$

$$0\le P,E,T_r,G$$

Here G is the number of scholarship/assistantship positions used by the group. The parameter t_g represents the fraction of internally funded staff time that needs to be spent on commitments other than the scientist's research (e.g. teaching). Similar to the previous problems, the Lagrangian for this problem is formed and Karush-Kuhn-Tucker conditions applied, leading to the following feasible solutions:

Conditions	Equipment (E)	Scientist's Research Execution Time (T_r)	Number of PhD Students Funded by Research Money (P)	Number of PhD Students Funded by Scholarships and Assistantships (G)
$r_p\le t_a+t_f$ & $r_p\le t_a/(1-t_g)$	$(1-\alpha)/t_e$	α	0	0
$r_p\le t_a+t_f$ & $r_p>t_a/(1-t_g)$ &	$\frac{(1-\alpha)}{t_e.\alpha}(T_r+r_pg(1-t_g))$	$\alpha.(1-gt_a-\frac{(1-\alpha)}{\alpha}r_pg(1-t_g))$	0	g

$1 > gt_a$				
$r_p \le t_a + t_f$ & $r_p > t_a/(1-t_g)$ & $1 \le gt_a$	$(1-\alpha)/t_e$	0	0	α / t_a
$r_p > t_a + t_f$ & $r_p \le t_a/(1-t_g)$	$(1-\alpha)/t_e$	0	$\frac{\alpha}{t_a + t_f}$	0
$r_p > t_a + t_f$ & $r_p > t_a/(1-t_g)$	$\frac{(1-\alpha)}{t_e.\alpha}(P + g(1-t_g))(t_f + t_a)$	0	$\frac{\alpha(1-gt_a) - g(1-t_g)(t_f+t_a)(1-\alpha)}{(t_f+t_a)}$	g

C-Simulating evolving practices in a community of scientists

To explore the relationship between the analytical model's predictions and the evolutionary dynamics of a behaviorally more realistic community of simulated scientists, we developed an agent-based model with the following features:

- One hundred scientists are modeled to be interacting over 1000 periods with each other, occasionally changing their research organization.
- Scientists are subject to the same production function as described in the basic model and are all exposed to the same underlying parameters (in the base-case: t_e=0.1, e_s=2, α=0.8; t_a=0.1, t_f=0.1; r_p is changed in reported scenarios).
- Scientists' research organization decisions include P (number of staff), E (number of equipment), and T_r (fraction of time spent on research). Below we call an instance of these three decisions as an 'organization'.
- Scientists are unaware of underlying parameters that shape their production function; they also do not follow any formal optimization process. However, they change their organizational decisions following these rules:
    - Every period, if a change was made in the previous period and the publication output declined, the scientist would revert back to the previous organization.
    - Each period with a probability that scales with O_{max}-O (O is the current output of the scientist and O_{max} the current maximum in the community), the scientist attempts a change in her organization.
    - If her current output is above community average, this change will be a random walk in the vicinity of the current organization. If her current output is below community average, she will switch to following the community norm on the three decisions shaping the organization.
    - Community norm is defined as the average on each decision across all members of the community.
- Simulations start with each scientist adopting a random set of decisions intendent of others in the community.

This simulation model is implemented in Vensim ™ and available as part of the online appendix, providing full documentation on remaining parameter values and implementation details. Interested readers can open the model using the free Vensim model reader and both explore model formulations and conduct simulations with different parameter settings.

The purpose of this simulation model is limited to illustrating the basic insight that the simple analytical model we introduce in the paper can be informative about the trajectory of more complex evolutionary dynamics in a scientific community. Thus full analysis of this simulation model and exploring various alternative formulations is beyond the scope of the current paper. Here we report two simulation runs for how the community norms on the three decisions comprising the organization of scientific work evolve over time. The simulation runs show the optimal solutions for P, T_r, and E (from the analytical model) along with the evolving norms for each. The graph on the right shows the case with r_p=0.l, leading to dominance of the hands-on model. The graph on the left simulates the community with r_p=0.5, i.e. a setting where funded model dominates.

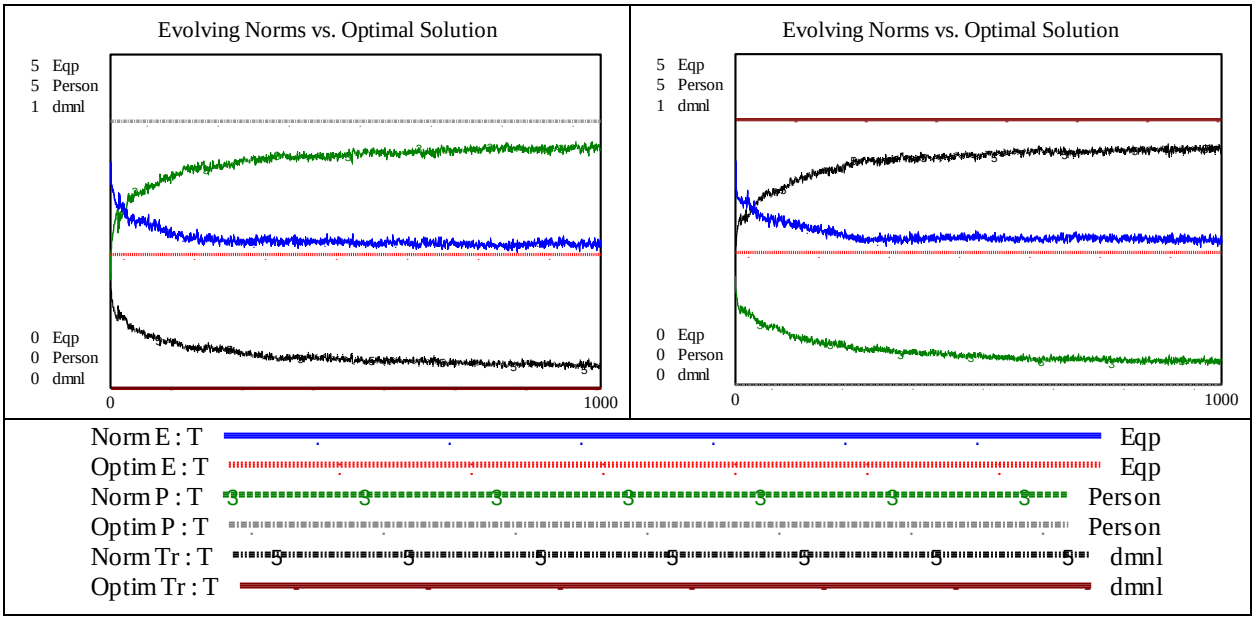


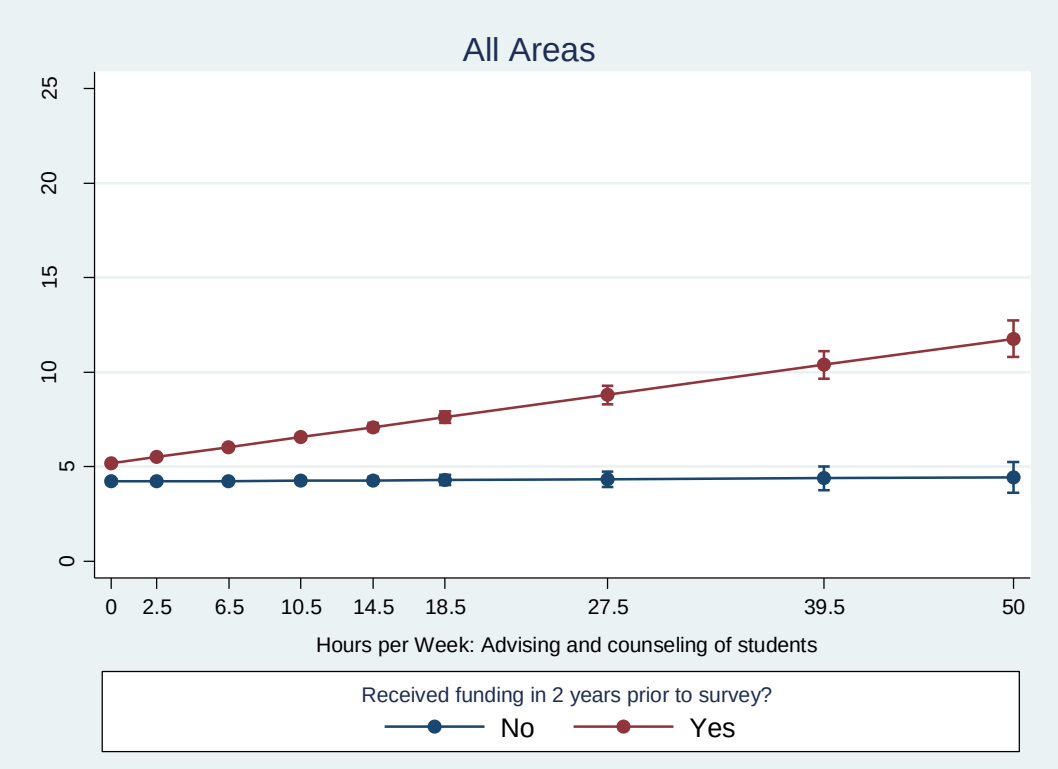
Figure S1- Evolving norms for equipment levels (E), number of staff (P) and research time (T_r) in a community of 100 scientists. Left: a setting where funded model dominate. Right: a setting where hands-on model dominates.

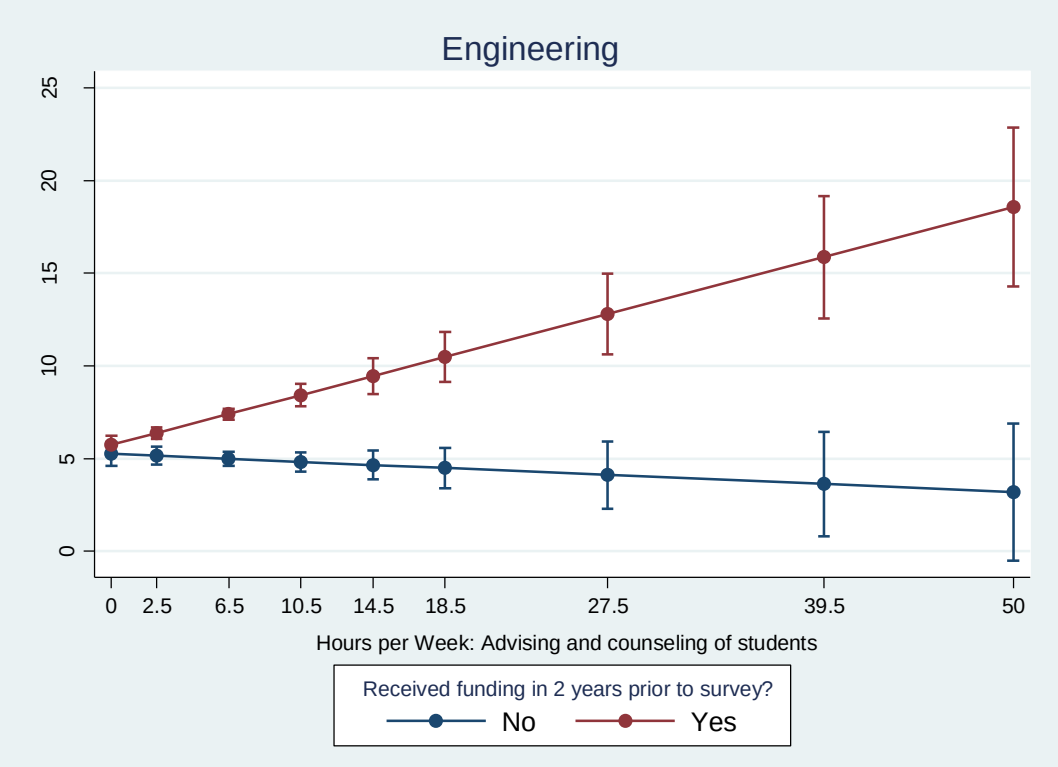
In both cases The community norm over time converges towards the optimal solution, even though none of the simulated scientists are explicitly maximizing their publications. The convergence is very fast initially, but slows down as random exploration offers few improvement opportunities when most scientists are close to the optimal organization, and incentives for changing the organization decline as laggards catch up to the more productive members of the community.

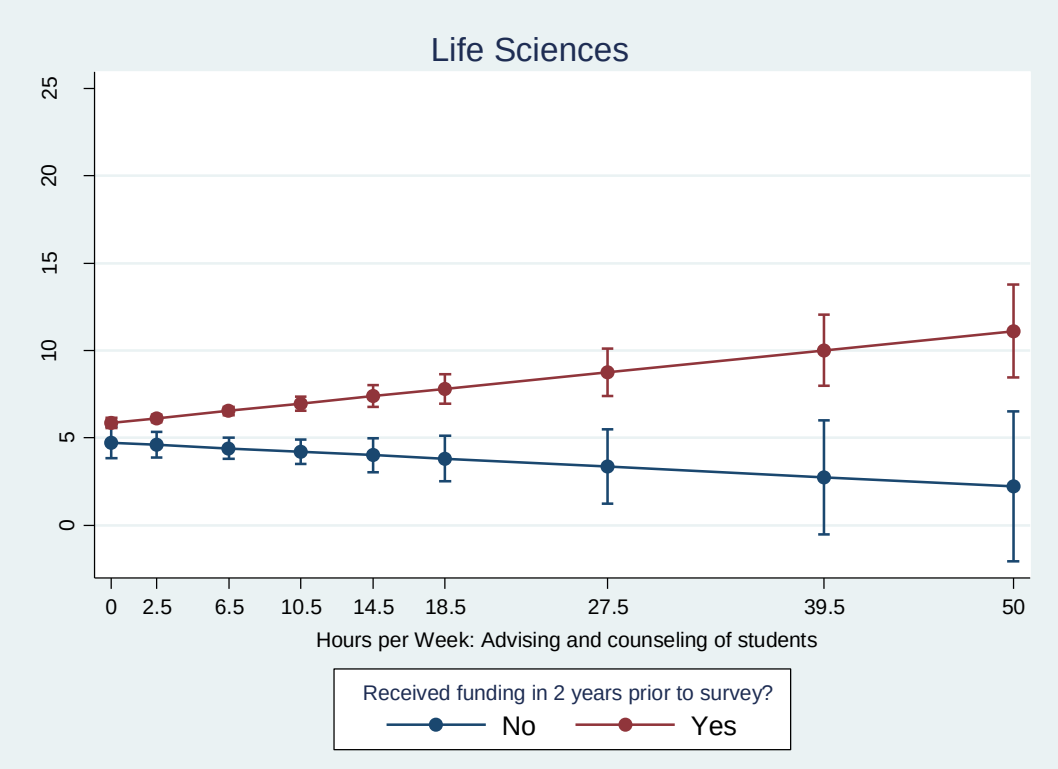
D-Funding and student advising interaction graphs

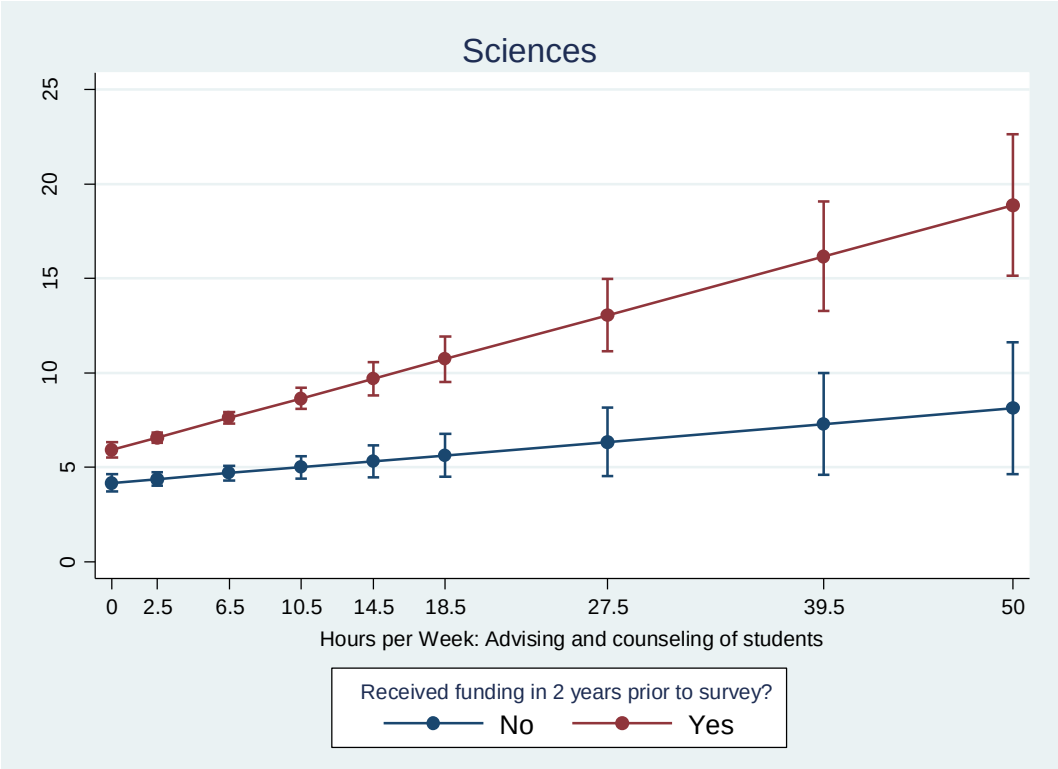
The eight graphs below further depict the relationship between time spent advising/counseling students and predicted research output in the presence and absence of funding across all areas and within each area separately. The graphs for engineering, life sciences, sciences, and psychology show a clear increase in research output as funded faculty spend more time advising/counseling students. Few research productivity benefits accrue, however, from allocating time to advising/counseling students in the absence of funding in these areas. In comparison, in social sciences (excluding psychology), humanities, and management, spending more time with students has little effect on research output regardless of whether faculty received funding over the 2 years prior to each survey.

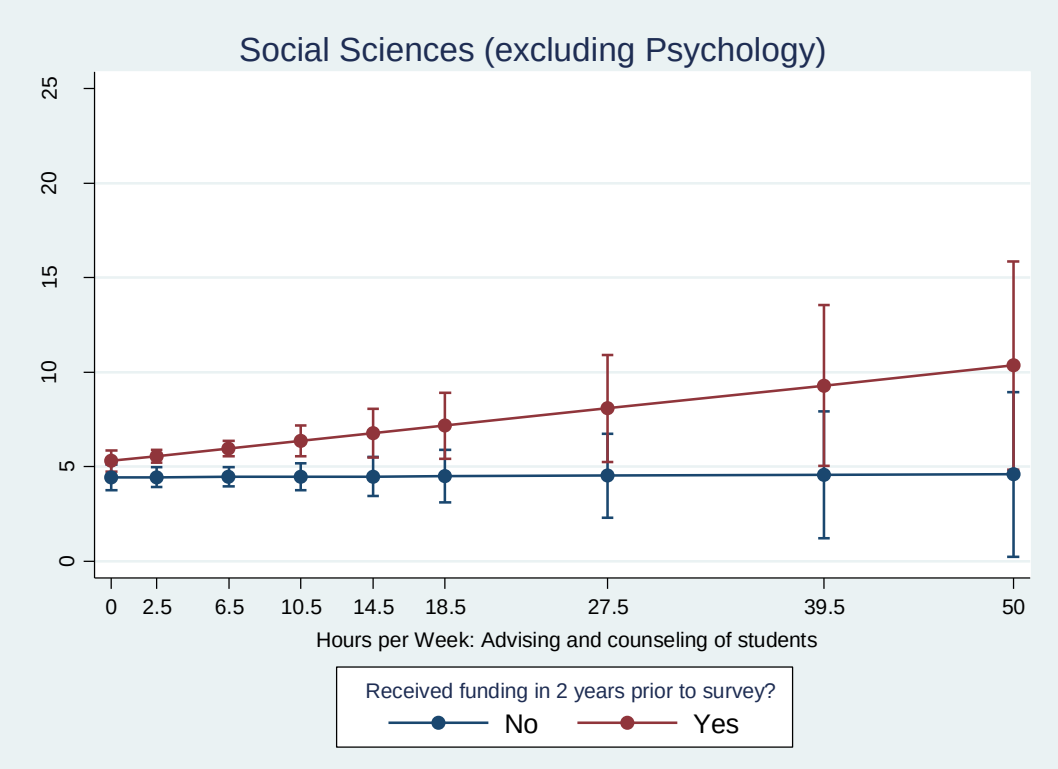
Figure S2- The interaction between funding, advising/counseling students, and research output across major areas (bars on graphs indicate 95% confidence intervals)

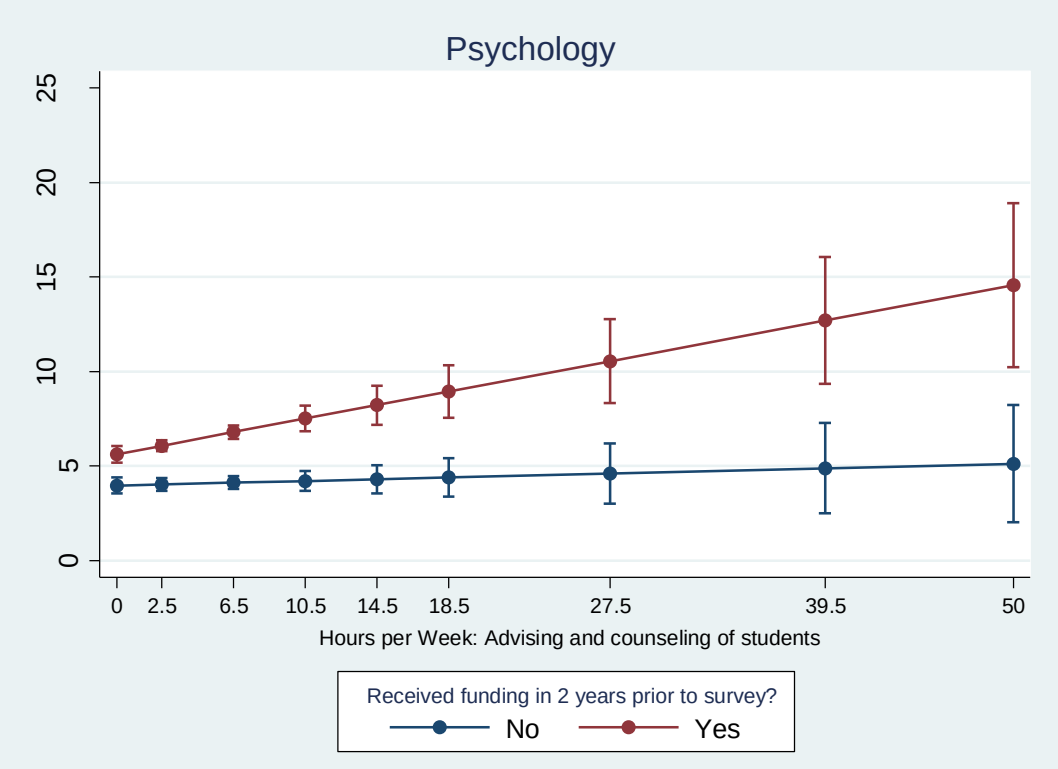

All Areas
0
5
10
15
20
25
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes

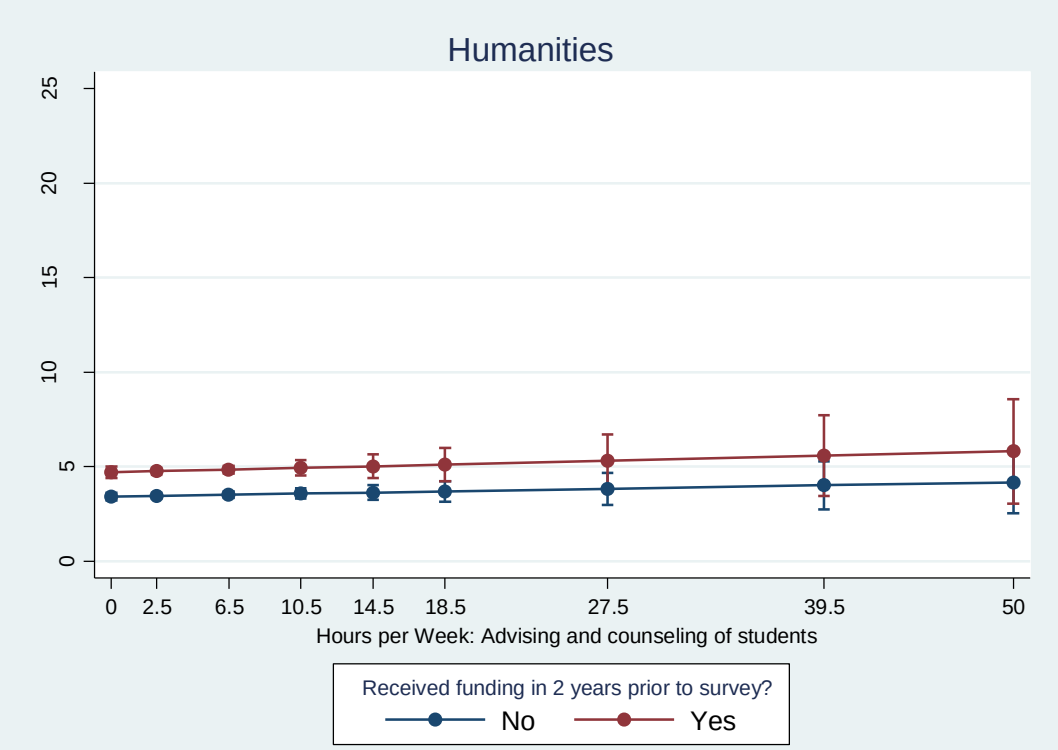

Engineering
0
5
10
15
20
25
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes

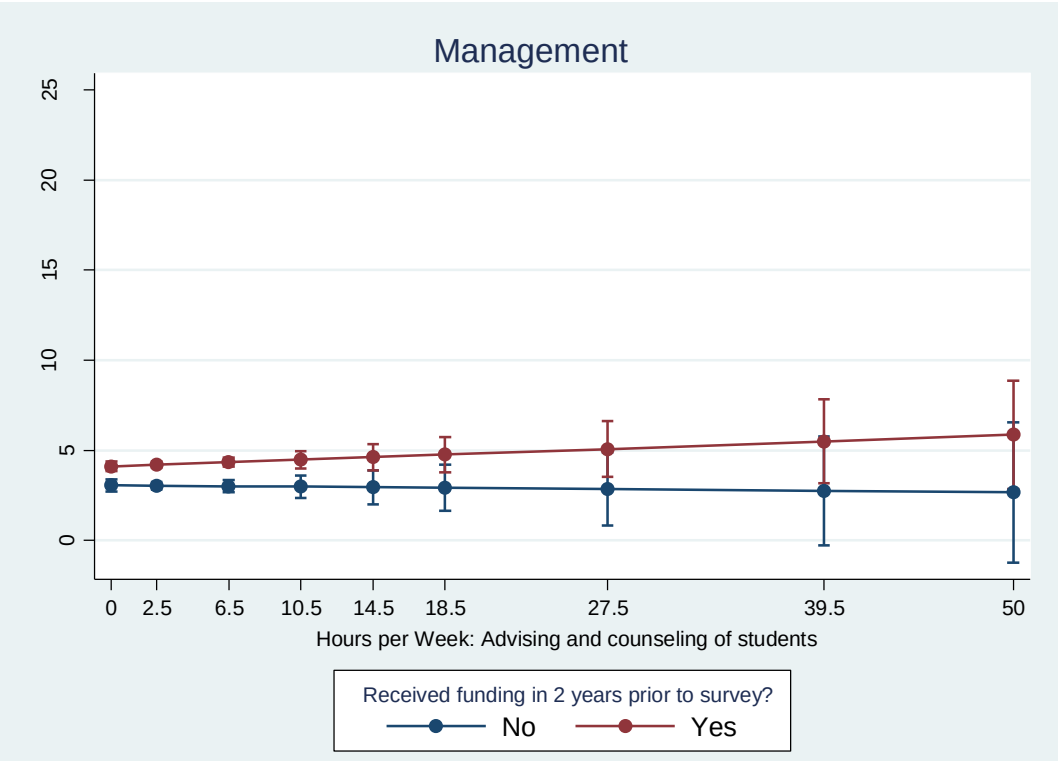

Life Sciences
0
5
10
15
20
25
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes


Sciences
0
5
10
15
20
25
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes


Social Sciences (excluding Psychology)
0
5
10
15
20
25
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes


Psychology
0
5
10
15
20
25
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes


Humanities
25
20
15
10
5
0
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes


Management
25
20
15
10
5
0
0
2.5
6.5
10.5
14.5
18.5
27.5
39.5
50
Hours per Week: Advising and counseling of students
Received funding in 2 years prior to survey?
No
Yes

E-Survey questions

The survey below was distributed to all faculty at the Massachusetts Institute of Technology and Virginia Tech. If a scientist wanted to report more than one incident, the questions on block 2 were repeated for each incident.

BLOCK 1

Your First Name _____

Your Last Name _____

On the following page please share instances in the past 20 years (1996-2016) when one or more of your direct advisees (a Ph.D. student, post-doc, or lab technician who was working under your supervision) went on leave/departed for an extended period of time (more than 2 months), or joined the team significantly later than expected, due to unforeseeable events (e.g. visa delays, family commitments, health problems).

How many instances would you like to report (please limit to the most notable 3 instances)

BLOCK 2

Consider the instance when one of your advisees/research staff was absent unexpectedly and answer these questions based on that instance.

In what year this instance started?

▼ 1996 ... 2016

Starting Month (mark the closest month you remember)

▼ January ... December, Can't Remember

How long did this absence last?

- ○ 2-4 Months
- ○ 5-8 Months
- ○ 9-12 Months
- ○ More than 12 Months
- ○ Can't Remember

Advisee type

- ○ early stage PhD student
- ○ late stage PhD student
- ○ post-doc associate
- ○ technician or lab staff
- ○ Other _____

Advisee gender

- ○ Male
- ○ Female

Advisee primary source of funding at the time

- ○ Research funding I supplied
- ○ Teaching assistantship
- ○ Fellowships or other funding
- ○ No Funding/ Self-funded
- ○ Can't Remember

What is your subjective evaluation of the quality, research wise, of that advisee compared to her/his peers?

- ○ Below average
- ○ Typical
- ○ Above average

Reason for absence

- ○ Visa or immigration issues
- ○ Health
- ○ Family
- ○ Other/Unknown

Do you think the underlying reason for absence impacted your advisee's performance before their absence?

- ○ No
- ○ Probably not
- ○ Maybe
- ○ Yes

How many active PhD students, post-docs, and other research staff did you advise and manage at the time?

- ○ Fewer than 3
- ○ 3-5
- ○ 6-10
- ○ More than 10

Approximately how much research funding did you have at the time on an annual basis?

- ○ Under $20,000
- ○ $20k-50k
- ○ $50k-100k
- ○ $100-200k
- ○ $200-500k
- ○ More than $500k

What was your rank at the time?

- ○ Assistant Professor
- ○ Associate Professor
- ○ Full Professor
- ○ Other

Your institution at the time of this instance?

- ○ MIT
- ○ Virginia Tech
- ○ Other