

Supplement to:

Azoulay, Pierre, and Freda B. Lynn. 2020. "Self-Citation, Cumulative Advantage, and Gender Inequality in Science." *Sociological Science* 7: 152-186.

Figure A1: Distribution of relatedness rank for 1,303,113 related citing/cited pairs

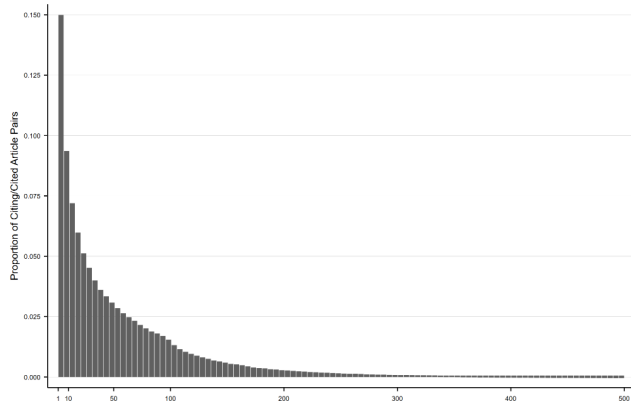


Table A1: Scientists' Initial Positions

	Men	Women	Total
Tenure-track Academic	1,996 (76.39)	682 (64.71)	2,678 (73.03)
Staff Scientist/Research Associate	124 (4.75)	91 (8.63)	215 (5.86)
Academic Administration	7 (0.27)	7 (0.66)	14 (0.38)
Science/Health Policy	5 (0.19)	5 (0.47)	10 (0.27)
Higher Ed Adjunct Teaching	10 (0.38)	15 (1.42)	25 (0.68)
Clinical Medicine	28 (1.07)	11 (1.04)	39 (1.06)
High-school Science Teaching	2 (0.08)	6 (0.57)	8 (0.22)
Big Pharma	176 (6.74)	77 (7.31)	253 (6.90)
Biotech	205 (7.85)	100 (9.49)	305 (8.32)
Patent Law	7 (0.27)	10 (0.95)	17 (0.46)
Venture Capital	11 (0.42)	1 (0.09)	12 (0.33)
Biopharma Consulting	9 (0.34)	3 (0.28)	12 (0.33)
Scientific Writing/Editing	4 (0.15)	10 (0.95)	14 (0.38)
Other Industry	12 (0.46)	10 (0.95)	22 (0.60)
Unknown	17 (0.65)	26 (2.47)	43 (1.17)
Total	2,613 (100.00)	1,054 (100.00)	3,667 (100.00)

Note: Column Percentages in Parentheses. Observations with "unknown" position type are not lost to follow up. These are individuals who not only exited science (as ascertained by the complete absence of any publication or patent), but have also seemingly dropped out of the labor force.

Table A2: Scientists' Last Positions

	Men	Women	Total
Tenure-track Academic	1,842 (70.49)	635 (60.25)	2,477 (67.55)
Staff Scientist/Research Associate	52 (1.99)	58 (5.50)	110 (3.00)
Academic Administration	59 (2.26)	45 (4.27)	104 (2.84)
Science/Health Policy	12 (0.46)	7 (0.66)	19 (0.52)
Higher Ed Adjunct Teaching	17 (0.65)	22 (2.09)	39 (1.06)
Clinical Medicine	38 (1.45)	20 (1.90)	58 (1.58)
High-school Science Teaching	7 (0.27)	13 (1.23)	20 (0.55)
Big Pharma	192 (7.35)	80 (7.59)	272 (7.42)
Biotech	227 (8.69)	63 (5.98)	290 (7.91)
Patent Law	19 (0.73)	19 (1.80)	38 (1.04)
Venture Capital	15 (0.57)	3 (0.28)	18 (0.49)
Biopharma Consulting	60 (2.30)	28 (2.66)	88 (2.40)
Scientific Writing/Editing	15 (0.57)	14 (1.33)	29 (0.79)
Other Industry	34 (1.30)	24 (2.28)	58 (1.58)
Unknown	24 (0.92)	23 (2.18)	47 (1.28)
Total	2,613 (100.00)	1,054 (100.00)	3,667 (100.00)

Note: Column Percentages in Parentheses. Observations with "unknown" position type are not lost to follow up. These are individuals who not only exited science (as ascertained by the complete absence of any publication or patent), but have also seemingly dropped out of the labor force.

Table A3: The effect of self-references on receipt of NIH R01 funding (OLS)

	(1)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
Woman Scientist	-0.02 ^{**} (0.01)	0.01 (0.01)		0.00 (0.02)		0.01 (0.01)		0.01 (0.02)	
Ln(Publications)		0.11 ^{**} (0.00)	0.06 ^{**} (0.01)	0.11 ^{**} (0.00)	0.06 ^{**} (0.01)	0.10 ^{**} (0.01)	0.07 ^{**} (0.01)	0.10 ^{**} (0.01)	0.07 ^{**} (0.01)
Ln(Publications) × Woman				0.00 (0.01)	-0.00 (0.01)			-0.00 (0.01)	-0.01 (0.01)
Fraction Self-References						0.44 ^{**} (0.09)	-0.16 [*] (0.09)	0.42 ^{**} (0.09)	-0.22 [*] (0.10)
Fraction Self-References × Woman								0.10 (0.20)	0.38 (0.27)
Scientist Fixed Effects	Included	Excluded	Included	Excluded	Included	Excluded	Included	Excluded	Included
Mean of Dependent Variable	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Adjusted R ²	0.01	0.05	0.14	0.05	0.14	0.05	0.14	0.05	0.14
Nb. of Scientists	2,678	2,678	2,673	2,678	2,673	2,678	2,673	2,678	2,673
Nb. of Scientist-Year Obs.	61,151	61,151	61,146	61,151	61,146	61,151	61,146	61,151	61,146

Note: Linear Probability Model (OLS) estimates. An observation corresponds to a scientist-year. The dependent variable is an indicator variable equal to one if a scientist is awarded an R01 grant from the NIH in the focal year, zero otherwise (it is exceedingly rare for a scientist to receive more than one such grant in a given year). Since scientists are only at risk of receiving an R01 grant if they are principal investigators in academic positions, we restrict the sample to scientists who begin their career in academia, from the start of their independent career, with the last observation corresponding to the year of exit from science, retirement, death, or 2015, whichever comes earlier. The time-varying covariates are lagged one year. The specifications also include degree indicator variables (MD and MD/PhD, PhD is the omitted category), scientist career age and its square, a full suite of calendar year indicator variables, and eight indicator variables corresponding to years of highest degree (in five year increments). Standard errors in parentheses, triple clustered at the investigator level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: The effect of self-references on the flow of publications (OLS)

	(1)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
Woman Scientist	-0.24 ^{**} (0.02)	-0.05 ^{**} (0.01)		0.10 ^{**} (0.04)		-0.05 ^{**} (0.01)		0.14 ^{**} (0.04)	
Ln(Publications)		0.60 ^{**} (0.01)	0.31 ^{**} (0.01)	0.62 ^{**} (0.01)	0.32 ^{**} (0.01)	0.56 ^{**} (0.01)	0.32 ^{**} (0.01)	0.58 ^{**} (0.01)	0.33 ^{**} (0.01)
Ln(Publications) × Woman				-0.05 ^{**} (0.01)	-0.04 ^{**} (0.01)			-0.08 ^{**} (0.02)	-0.06 ^{**} (0.02)
Fraction Self-References						1.43 ^{**} (0.22)	-0.16 (0.26)	1.14 ^{**} (0.24)	-0.37 (0.29)
Fraction Self-References × Woman								1.47 ^{**} (0.52)	1.09 [*] (0.60)
Scientist Fixed Effects	Included	Excluded	Included	Excluded	Included	Excluded	Included	Excluded	Included
Mean of Dependent Variable	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Adjusted R ²	0.05	0.33	0.46	0.33	0.46	0.33	0.46	0.33	0.46
Nb. of Scientists	3,667	3,667	3,661	3,667	3,661	3,667	3,661	3,667	3,661
Nb. of Scientist-Year Obs.	94,315	94,315	94,308	94,315	94,308	94,315	94,308	94,315	94,308

Note: OLS estimates. The dependent variable is the flow of original articles published in year t by the focal scientist, transformed using the inverse hyperbolic sine transformation, so that coefficient estimates are approximately interpretable as elasticities. An observation corresponds to a scientist-year, with the first observation being the year following the receipt of highest degree, and the last observation the year of exit from science, retirement, death or 2015, whichever comes earlier. The time-varying covariates are lagged one year. The specifications also include degree indicator variables (MD and MD/PhD, PhD is the omitted category), scientist career age and its square, a full suite of calendar year indicator variables, and eight indicator variables corresponding to years of highest degree (in five year increments). Standard errors in parentheses, triple clustered at the investigator level. ^{*} $p < 0.10$, ^{**} $p < 0.05$, ^{***} $p < 0.01$.