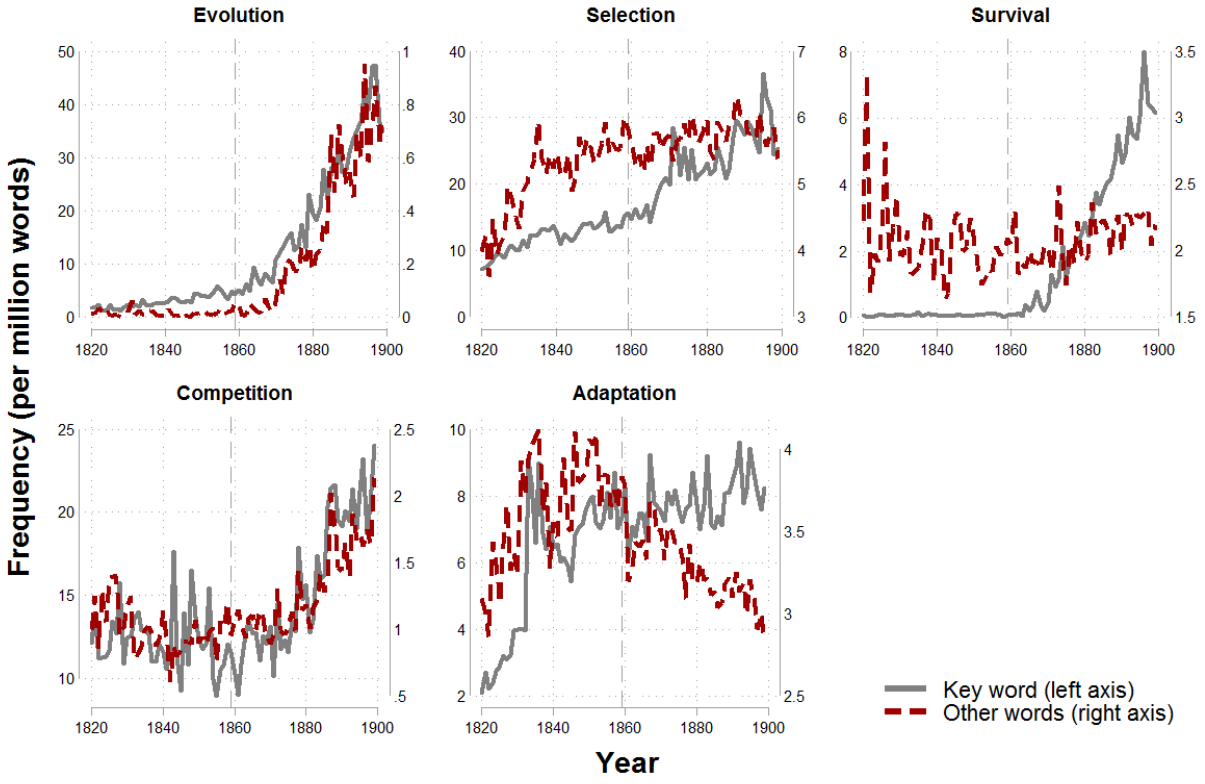
How Does Scientific Progress Affect Cultural Changes? A Digital Text Analysis

Michela Giorcelli, University of California – Los Angeles and NBER
Nicola Lacetera, University of Toronto and NBER
Astrid Marinoni, Georgia Institute of Technology

ONLINE APPENDIX: ADDITIONAL FIGURES AND TABLES

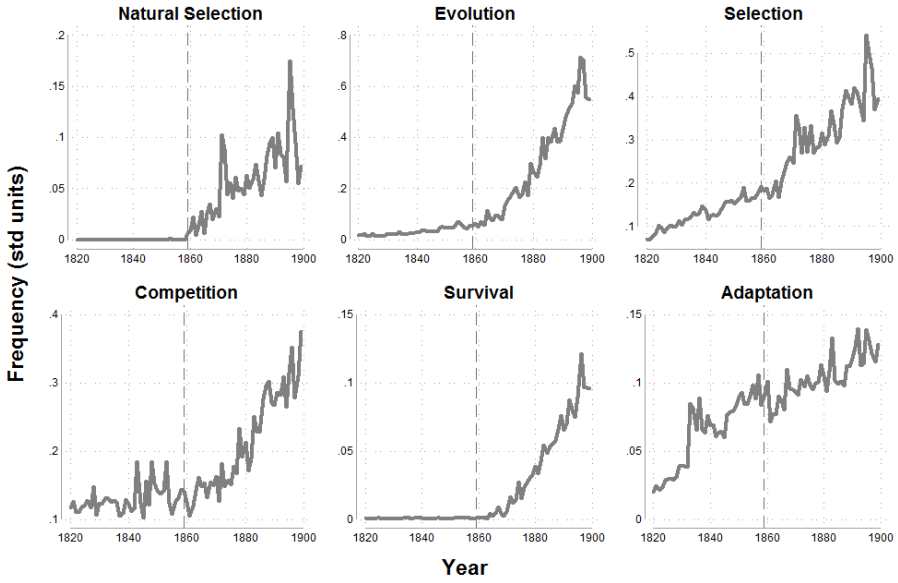
Figure A1: Frequencies (per 1 Million Words) of Key Darwinian Words and Other Words from the Same Roots in the Google Books Corpora



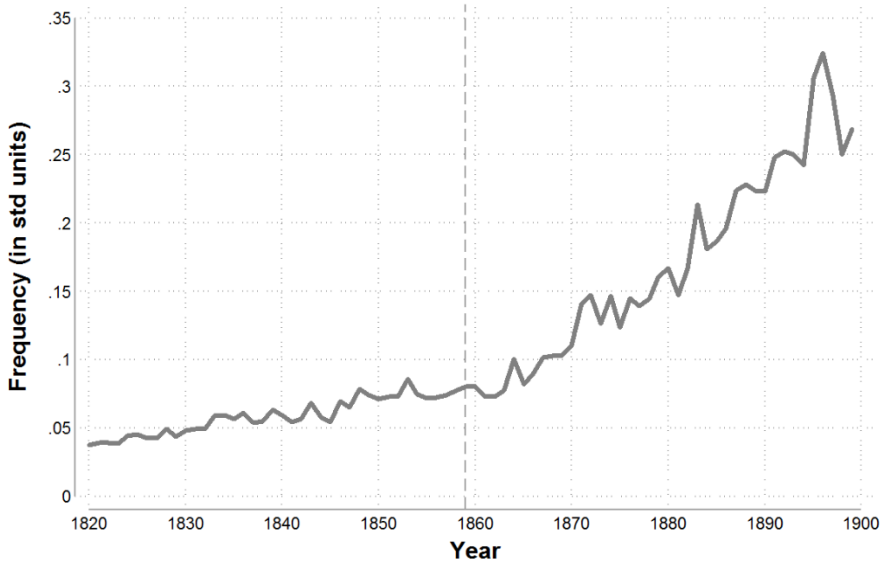
Notes: For each year, the graphs show the number of occurrences per one million words. The gray solid lines display the yearly frequency of each of the key Darwinian 1-grams as reported in the title of the chart; red dashed lines indicate the average frequency, per million words, of other words with the same roots of the key term. These are words are: adapted, adaptive, adapting, adapt, adaptability, adaptable, adapter, adaption, adaptogen, adaptogenic, adaptometer, adaptor, competitive, competitor, competitory, evolutionary, evolutionist, evolutive, evolute, select, selected, selecting, selecta, selectance, selectee, selective, selectivity, selectman, selector, selectorate, survive, survivor, survivable, survivalist, survivorship.

Figure A2: Frequencies Darwinian in the Google Books Corpus, as a Share of the Standard Deviation of Frequencies of All Words in Each Given Year

Panel A: Term by Term

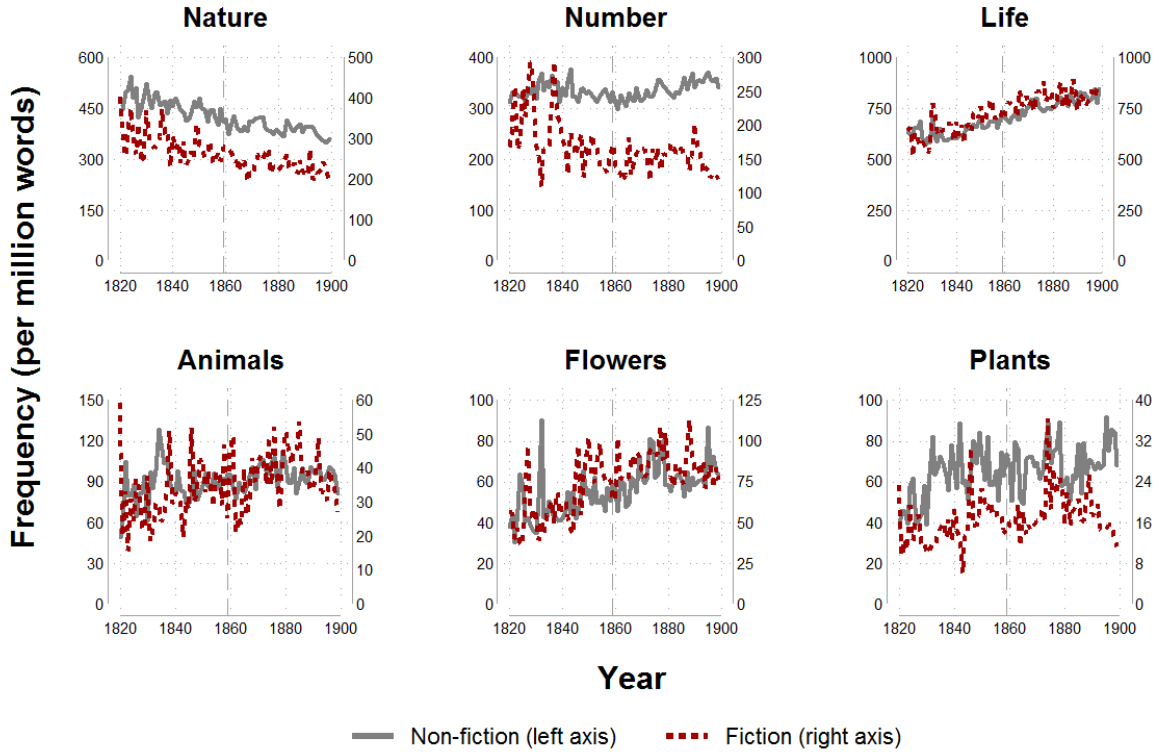


Panel B: Average Frequency across Six Terms



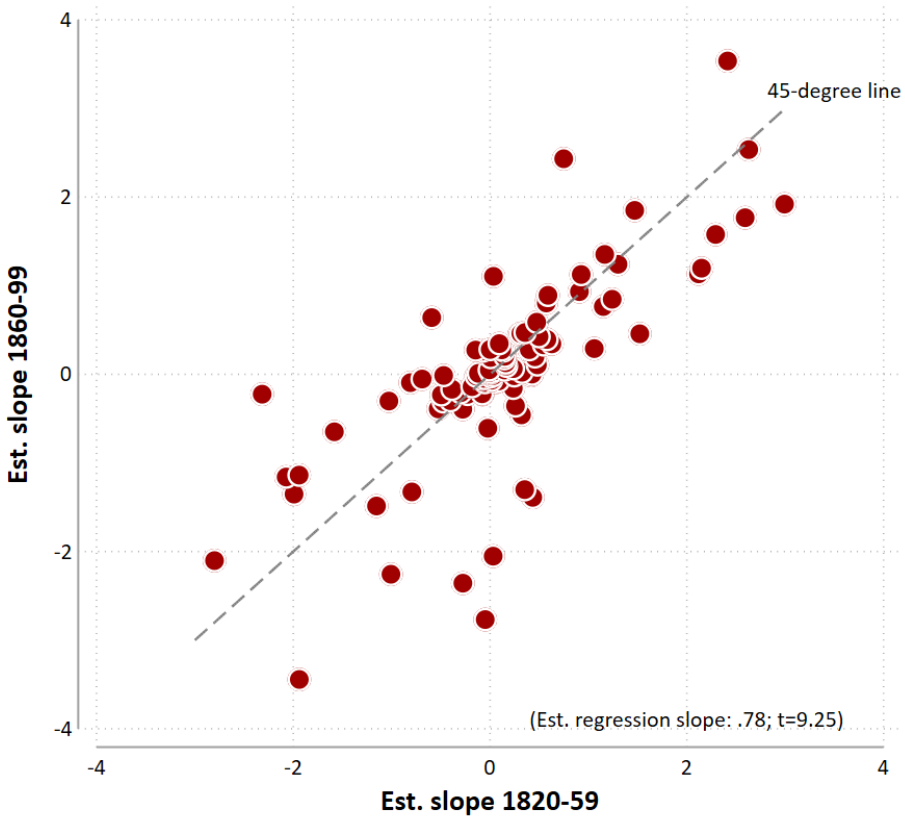
Notes: For each year, the graphs show the number of occurrences of the word or phrase reported on top (panel A) and the average across the six frequencies (panel B), as shares of the standard deviation of the frequencies of all one-grams present in the Google Books corpus in each year.

Figure A3: Frequencies (per One Million Words) of Select Generic Terms in the Google Books Corpora



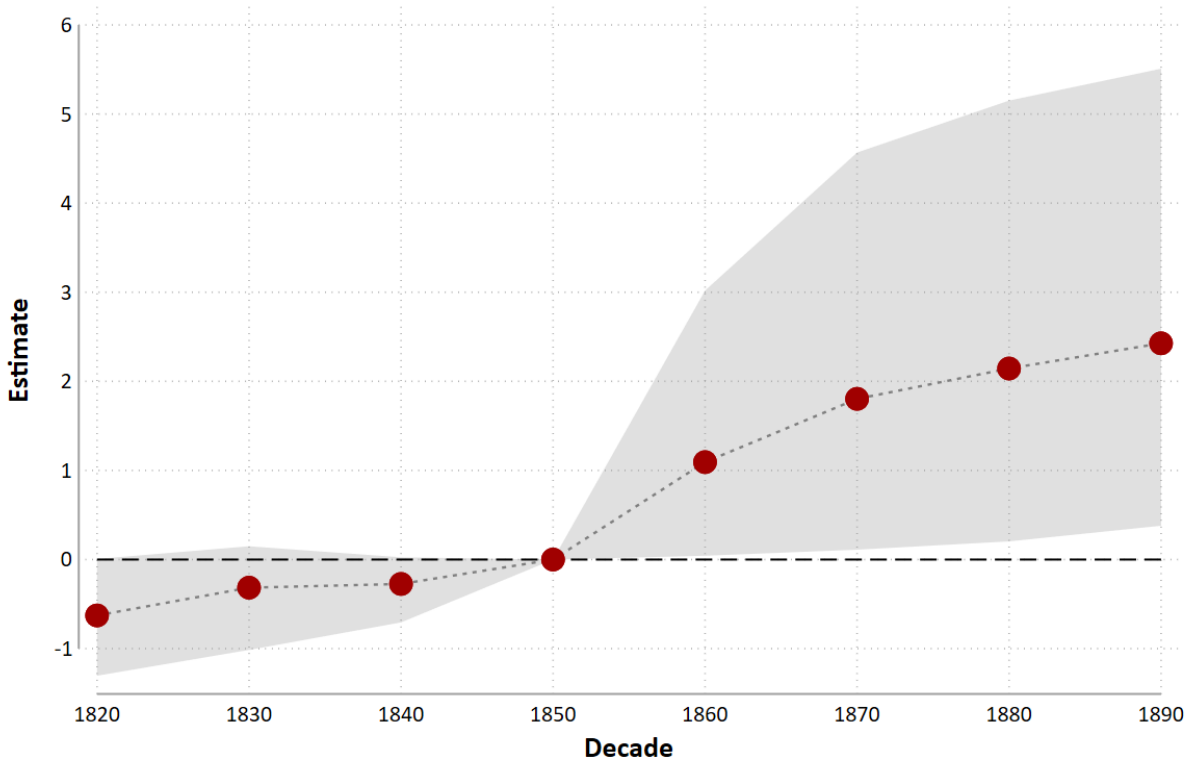
Notes: For each year, the graphs show the number of occurrences of the word or phrase reported on top per one million words. In panel A, the gray solid line displays the yearly frequency, whereas the red dashed line is a median band plot with 8 intervals (each of ten years). Note that also the denominators for the calculation of the relative frequencies are separate for fiction and non-fiction.

Figure A4: Correlation in the Estimates of Slopes in the Frequency of 99 High-Frequency (Control) Words in the 1820-59 and 1860-99 Period



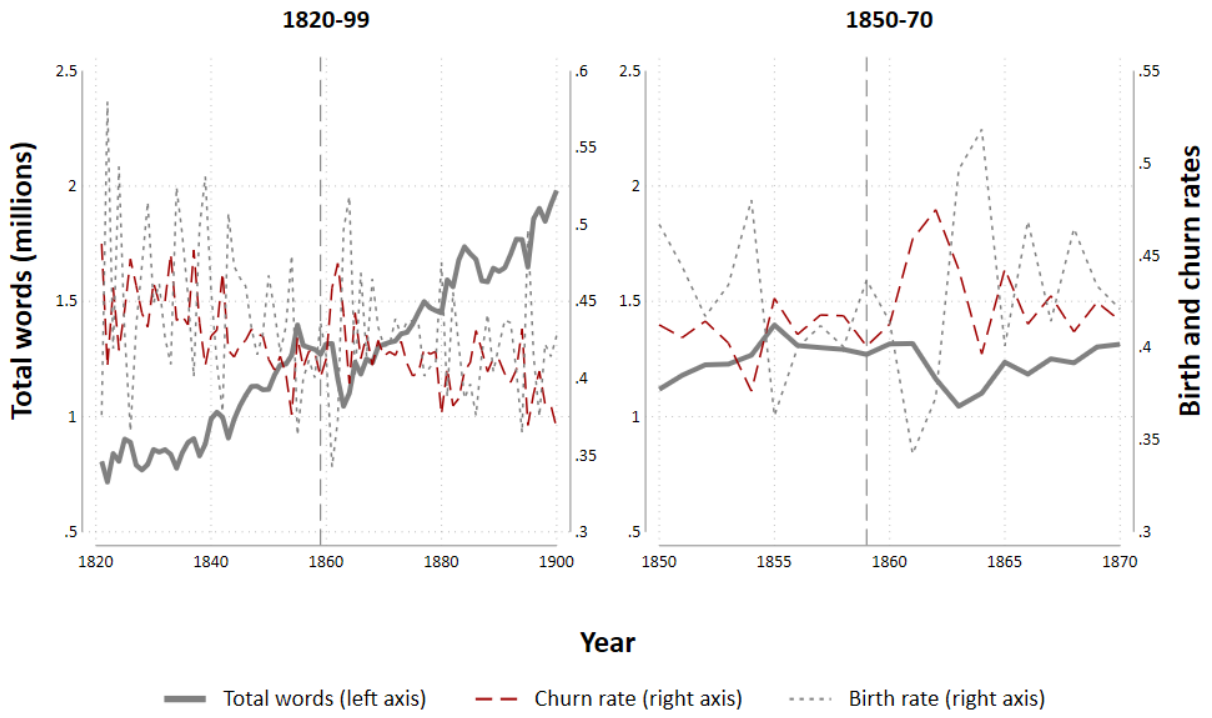
Notes: Each dot represent the OLS estimates $(\hat{\beta}_{1w}, \hat{\beta}_{2w})$ from regressions, for each word w: $y_{wt} = \alpha_w + \beta_{1w} \min(t, 59) + \beta_{2w} (\max(t, 59) - 59) + \varepsilon_{wt}$, where y_{wt} is the frequency per one million word for each yeat t=20, 21, …99. The pair $(\hat{\beta}_{1w}, \hat{\beta}_{2w})$ thus reporesents the estimates of the slope of the diffusion of each word in the 1820-59 period (x-axis) and the 1860-99 period (y-axis). The estimated regression slope and t-stat reported in the chart refer to the OLS estimate $\hat{\delta}$ from the following equations, where each observation is a given word: $\hat{\beta}_{2w} = \gamma_w + \delta_w \hat{\beta}_{1w} + \eta_w$.

Figure A5: Differences-in-Differences Estimates of the Average Frequency of Darwinian and Generic Concepts in Each Decade between 1820 and 1899: Alternative Control Sample



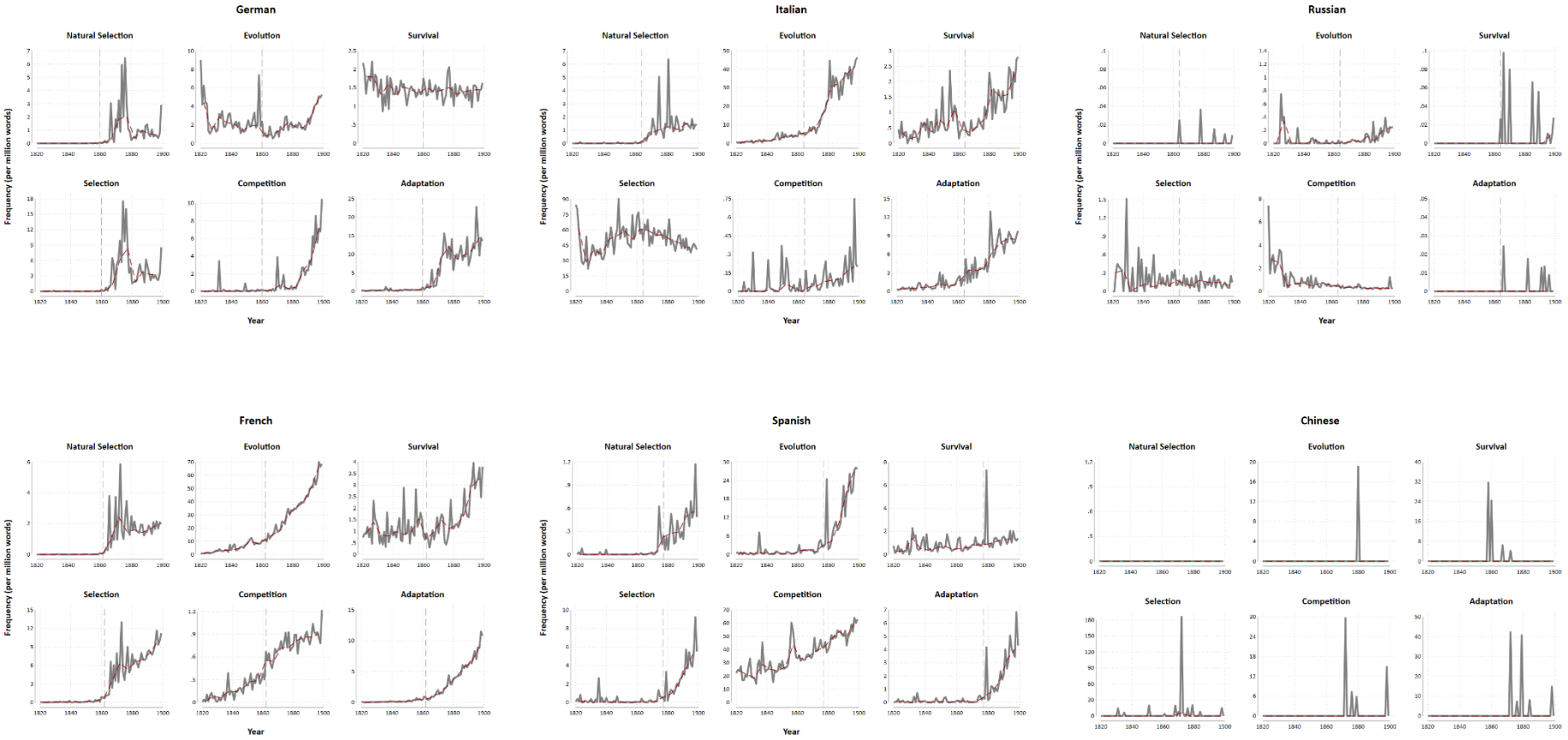
Notes: Each dot in the graph represents the estimate of the parameters δ_j from the following regression model: $\ln(y_{wt}) = \alpha_w + \beta_w \mathbf{1}(Darwinian) + \sum_{j=2}^{4} \gamma_i \mathbf{1}(j0 \leq t \leq j9) + \sum_{j=6}^{9} \gamma_j \mathbf{1}(j0 \leq t \leq j9) + \sum_{i=2}^{4} \delta_i \mathbf{1}(j0 \leq t \leq j9) * \mathbf{1}(Darwinian) + \sum_{j=6}^{9} \delta_j \mathbf{1}(j0 \leq t \leq j9) * \mathbf{1}(Darwinian) + \varepsilon_{wt}$, where y_{wt} is the frequency of use of a word per million words used and the omitted (or baseline) decade is 1850-59. The control sample is composed by the 100 words whose frequency between 1855 and 1858, the years immediately before the publication of *On the Origin of Species*, was closer in absolute value to the average frequency of the six Darwinian terms. Because the observed frequency is equal to zero in some cases, we add 0.001 to each frequency (0.001 is half of the lowest positive frequency per million words in our sample). The shaded area represents 95% confidence intervals that we computed using a wild bootstrap procedure (MacKinnnon and Webb 2019). Results are almost identical if we use an arcsine: $z = \ln\left(y + \sqrt{y^2 + 1}\right)$, or if we apply the GMM procedure described in Bellego and Pape (2019) to estimate the parameters (the confidence intervals from bootstrapped standard errors are narrower in this last case).

Figure A6: Total Number of Words in the Google Books Corpus per Year, and Rates of Entry and Exit of Words.



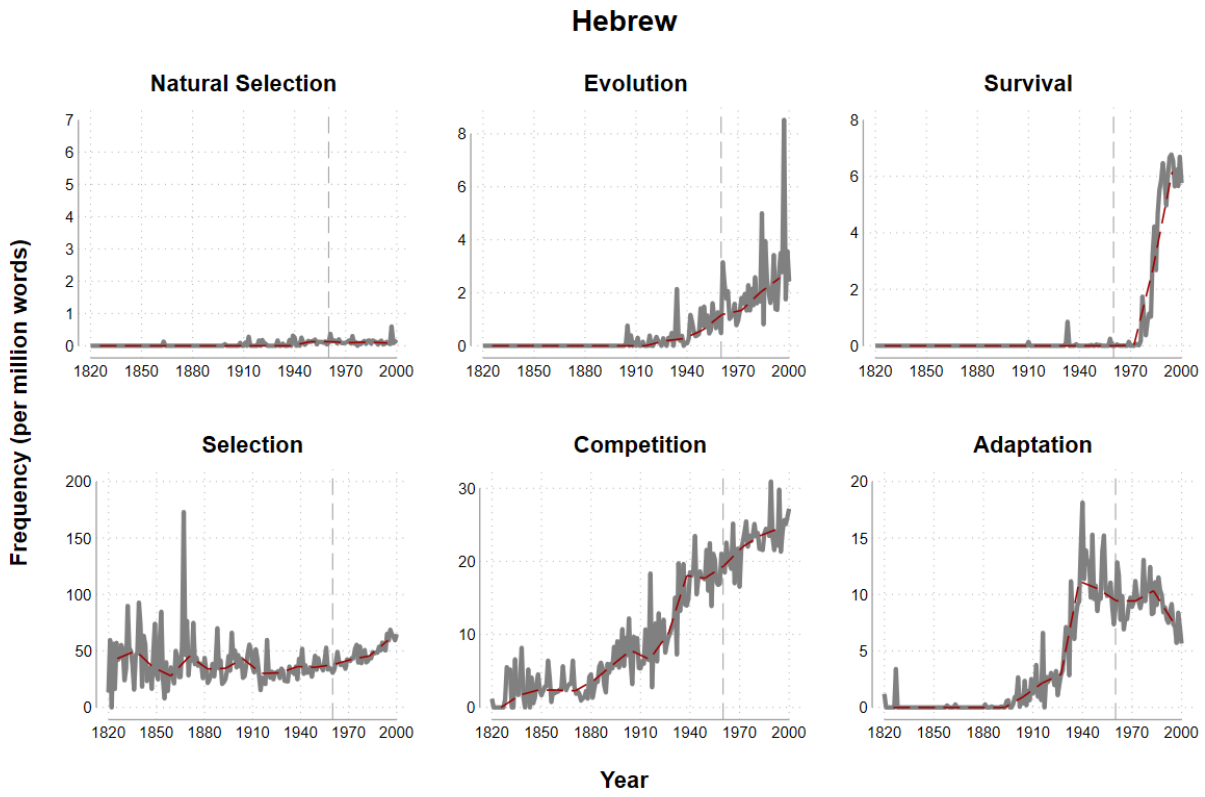
Notes: The graphs report the total number of distinct words (1-grams) present in the Google Books corpus in each year between 1820 and 1899, and the share of words that number of words that "entered" the corpus in a given year, i.e. they were present in year t but not in year t-1 (birth rate), and those that "exited", i.e. they were present in year t-1 but not in year t (churn rate).

Figure A7: Frequencies (per One Million Words) of Darwinian Words in Six Languages Other than English



Notes: For each year, the figures report the number of occurrences (per million words) in the language indicated on top of a graph; the vertical dashed lines are in correspondence of the year of the first published translation of *On the Origin of Species* in a given language.

Figure A8: Frequencies (per One Million Words) of Darwinian Words in Hebrew



Notes: For each year, the figures report the number of occurrences (per million words) in the language indicated on top of a graph; the vertical dashed lines are in correspondence of the year of the first published English version of *On the Origin of Species*. The first translation in Hebrew was in 1960.

Figure A9: Semantic Associations: Additional Word Pairs

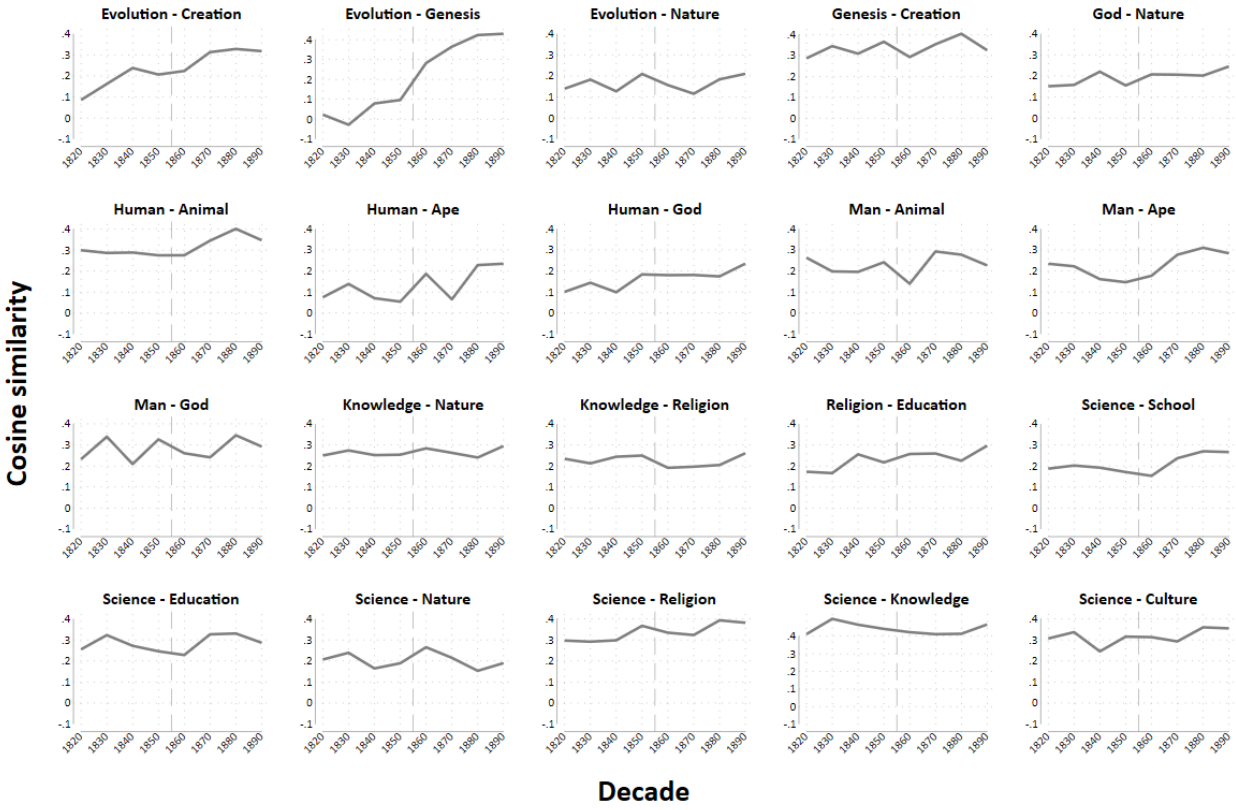


Table A1: Generic Words

action	forms	parent
advantage	genera	parts
animal	generations	period
animals	genus	periods
beings	group	plant
birds	groups	plants
breeds	habits	points
case	hand	pollen
cases	hybrids	power
change	importance	principle
changes	individuals	process
character	inhabitants	productions
characters	insects	reason
class	instance	respect
climate	instincts	sea
conditions	islands	seeds
country	kind	size
degree	kinds	species
descendants	land	state
descent	life	sterility
development	man	structure
difference	manner	subject
differences	means	tendency
difficulty	modification	theory
eggs	naturalists	time
fact	nature	variation
facts	number	variations
fertility	numbers	varieties
flower	offspring	variety
flowers	older	view
form	organ	water
formation	organization	world
formations	organs	years

Notes: The table lists the 99 most frequent nouns in *On the Origins of Species*, which we used as controls for the Darwinian concepts in some of the analyses.

Table A2: Spline Regression Analyses (eight knots) – Frequency of Darwinian Concepts

Word/phrase:	Natural Selection	Evolution	Selection	Competition	Survival	Adaptation
	(1)	(2)	(3)	(4)	(5)	(6)
Slope(1820-29)	-0.001	0.024	0.390***	0.149	0.003	0.289***
	(0.001)	(0.033)	(0.052)	(0.135)	(0.003)	(0.085)
Slope(1830-39)	0.001	0.076***	0.183***	-0.145	0.001	0.284***
	(0.003)	(0.024)	(0.063)	(0.117)	(0.002)	(0.057)
Slope(1840-49)	-0.001	0.135***	0.101	0.147	-0.001	-0.023
	(0.010)	(0.031)	(0.061)	(0.102)	(0.002)	(0.061)
Slope(1850-59)	0.005	0.053	-0.011	-0.264**	-0.003	0.051
	(0.039)	(0.052)	(0.085)	(0.113)	(0.005)	(0.051)
Slope(1860-69)	0.369***	0.438***	0.828***	0.174*	0.045***	0.017
	(0.109)	(0.093)	(0.159)	(0.090)	(0.014)	(0.046)
Slope(1870-79)	0.007	0.941***	0.008	0.168	0.190***	0.001
	(0.096)	(0.179)	(0.144)	(0.114)	(0.017)	(0.047)
Slope(1880-89)	0.240***	1.336***	0.579***	0.619***	0.253***	0.046
	(0.064)	(0.226)	(0.123)	(0.125)	(0.024)	(0.042)
Slope(1890-99)	-0.008	1.151**	0.129	0.140	0.192***	0.033
	(0.145)	(0.460)	(0.280)	(0.159)	(0.049)	(0.046)
Observations	80	80	80	80	80	80
R-squared	0.814	0.969	0.906	0.792	0.982	0.773

Notes: For each word and phrase, the estimates refer to the slope of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with eight knots at 19, 29, …, 89. The regression equation for a given word w is: $y_{wt} = \alpha_w + \beta_{1w}\min\{t, 29\} + \beta_{2w}(Max\{\min\{t, 39\}, 29\} - 29) + \beta_{3w}(Max\{\min\{t, 49\}, 39\} - 39) + \beta_{4w}(Max\{\min\{t, 59\}, 49\} - 49) + \beta_{5w}(Max\{\min\{t, 69\}, 59\} - 99) + \beta_{6w}(Max\{\min\{t, 79\}, 69\} - 69) + \beta_{7w}(Max\{\min\{t, 89\}, 79\} - 79) + \beta_{8w}(Max\{\min\{t, 99\}, 89\} - 89) + \varepsilon_{wt}$. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

.

Table A3: Spline Regression Analyses (eight knots) – Frequency of Select Generic Concepts

Word:	Nature	Number	Life	Animals	Flowers	Plants
	(1)	(2)	(3)	(4)	(5)	(6)
1820-29	0.760	2.655**	-2.180	2.664*	0.803	1.523*
	(2.677)	(1.095)	(3.208)	(1.557)	(1.111)	(0.774)
1830-39	-3.178*	0.054	-1.842	0.260	-0.456	1.210
	(1.894)	(1.186)	(2.505)	(0.846)	(0.687)	(0.733)
1840-49	-0.509	-1.644*	7.510***	-1.190*	1.216***	0.162
	(1.269)	(0.978)	(1.028)	(0.695)	(0.338)	(0.617)
1850-59	-4.066***	-1.231	0.723	0.660	-0.480	-0.703
	(1.421)	(0.777)	(1.140)	(0.508)	(0.397)	(0.657)
1860-69	-0.656	0.351	7.011***	0.463	1.429***	0.631
	(1.255)	(0.722)	(1.475)	(0.455)	(0.432)	(0.527)
1870-79	-1.611	1.531**	0.978	0.469	0.187	-0.052
	(1.023)	(0.768)	(1.321)	(0.572)	(0.514)	(0.638)
1880-89	0.712	1.003	3.458**	-0.553	-0.837*	-0.380
	(0.771)	(0.811)	(1.357)	(0.532)	(0.500)	(0.657)
1890-99	-4.699***	-0.366	2.131	-0.258	1.182**	1.215*
	(0.680)	(0.919)	(2.232)	(0.613)	(0.472)	(0.725)
Observations	80	80	80	80	80	80
R-squared	0.792	0.400	0.901	0.264	0.515	0.349

Notes: For each word and phrase, the estimates refer to the slope of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with eight knots at 19, 29, …, 89. The regression equation for a given word w is: $y_{wt} = \alpha_w + \beta_{1w} \min\{t, 29\} + \beta_{2w}(Max\{\min\{t, 39\}, 29\} - 29) + \beta_{3w}(Max\{\min\{t, 49\}, 39\} - 39) + \beta_{4w}(Max\{\min\{t, 59\}, 49\} - 49) + \beta_{5w}(Max\{\min\{t, 69\}, 59\} - 99) + \beta_{6w}(Max\{\min\{t, 79\}, 69\} - 69) + \beta_{7w}(Max\{\min\{t, 89\}, 79\} - 79) + \beta_{8w}(Max\{\min\{t, 99\}, 89\} - 89) + \varepsilon_{wt}$. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Alternative Control Words

addresses	douglas	mutiny
alfred	edict	natural
ali	effusion	outrage
amusements	elephant	palestine
aperture	endeavours	pontiff
awe	enlargement	poor
banner	ether	populace
battalion	expiration	predecessor
beard	expulsion	presbyterian
bedford	farms	redress
beef	fees	refinement
believer	fore	rewards
boots	fragment	robes
borough	frederic	rotation
brook	grandson	score
burgundy	grievances	secrets
candle	hospitals	signature
canons	indictment	statues
caroline	inscriptions	superior
chances	insurgents	telescope
charcoal	interpreter	tie
clement	isabella	tombs
clock	jewels	torture
concession	lancaster	transfer
conjecture	lip	transverse
consistency	manifestations	tumor
contemporaries	marshall	tumult
correspondent	martyrs	universities
data	mathematics	validity
denomination	merchandise	vapour
diagnosis	meridian	vases
disaster	might	wolf
disciple	montrose	
domain	morgan	

Notes: The table lists the 100 words whose frequency between 1855 and 1858, the years immediately before the publication of Darwin's book, was closer in absolute value to the average frequency of our six Darwinian expressions.

Table A5: Spline Regression Analyses (one and eight knots) – Total words, Churn Rates and Exit Rates

Outcome variable:	Total unique words (millions)		Churn rate		Entry rate	
	(1)	(2)	(3)	(4)	(5)	(6)
Slope(1820-59)	0.012***		-0.001***		-0.001	
	(0.001)		(0.000)		(0.001)	
Slope(1860-99)	0.016***		-0.001***		-0.000	
	(0.001)		(0.000)		(0.000)	
Slope(1820-29)		0.002		0.001		-0.000
		(0.005)		(0.003)		(0.008)
Slope(1830-39)		0.007**		-0.001		0.001
		(0.003)		(0.001)		(0.003)
Slope(1840-49)		0.028***		-0.003**		-0.003
		(0.003)		(0.001)		(0.003)
Slope(1850-59)		0.011***		0.001		-0.003
		(0.004)		(0.001)		(0.002)
Slope(1860-69)		-0.007*		0.000		0.004**
		(0.004)		(0.001)		(0.002)
Slope(1870-79)		0.033***		-0.003**		-0.003
		(0.005)		(0.001)		(0.002)
Slope(1880-89)		0.010***		0.002*		-0.000
		(0.004)		(0.001)		(0.002)
Slope(1890-99)		0.026***		-0.003***		0.000
		(0.004)		(0.001)		(0.001)
Average	1.26		0.42		0.43	
Observations	80	80	80	80	80	80
R-squared	0.937	0.966	0.439	0.519	0.126	0.170

Notes: The outcome variables in the regressions whose estimates are in this table are the total number of distinct words (1-grams) present in the Google Books corpus in each year between 1820 and 1899, and the share of words that number of words that "entered" the corpus in a given year, i.e. they were present in year t but not in year t-1 (birth rate), and those that "exited", i.e. they were present in year t-1 but not in year t (churn rate). In columns (1), (3) and (5), Slope(1820-59) and Slope(1860-99) refer to the slopes of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with one knot at 59. The regression equation for a given word *w* is: $y_{wt} = \alpha_w + \beta_{1w} \min(t, 59) + \beta_{2w} (\max(t, 59) - 59) + \varepsilon_{wt}$. In columns (2), (4) and (6), the estimates refer to the slope of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with eight knots at 19, 29, …, 89.
The regression equation for a given word *w* is: $y_{wt} = \alpha_w + \beta_{1w} \min\{t, 29\} + \beta_{2w}(Max\{\min\{t, 39\}, 29\} - 29) + \beta_{3w}(Max\{\min\{t, 49\}, 39\} - 39) + \beta_{4w}(Max\{\min\{t, 59\}, 49\} - 49) + \beta_{5w}(Max\{\min\{t, 69\}, 59\} - 99) + \beta_{6w}(Max\{\min\{t, 79\}, 69\} - 69) + \beta_{7w}(Max\{\min\{t, 89\}, 79\} - 79) + \beta_{8w}(Max\{\min\{t, 99\}, 89\} - 89) + \varepsilon_{wt}$.
Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Spline Regression Analyses (one knots) – Frequency of Occurrences of the Names Charles Darwin, Alfred Wallace, Robert Chambers and Jean-Baptiste Lamarck in the English Google Books Corpus

Name-Last name	(1)	Last name	(2)
Slope(1820-59)	0.0004	Slope(1820-59)	0.0447***
	(0.0010)		(0.0163)
Slope(1820-59) X Alfred Wallace	-0.0001	Slope(1820-59) X Wallace	-0.0057
	(0.0010)		(0.0255)
Slope(1820-59) X Robert Chambers	0.0019	Slope(1820-59) X Robert	0.1153***
	(0.0012)		(0.0277)
Slope(1820-59) X Jean Baptiste Lamarck	-0.0004	Slope(1820-59) X Lamarck	-0.0703***
	(0.0010)		(0.0191)
Slope(1860-99)	0.0218***	Slope(1860-99)	0.3516***
	(0.0031)		(0.0355)
Slope(1860-99) X Alfred Wallace	-0.0187***	Slope(1860-99) X Wallace	-0.2088***
	(0.0031)		(0.0419)
Slope(1860-99) X Robert Chambers	-0.0208***	Slope(1860-99) X Chambers	-0.3383***
	(0.0032)		(0.0419)
Slope(1860-99) X Jean Baptiste Lamarck	-0.0218***	Slope(1860-99) X Lamarck	-0.3404***
	(0.0031)		(0.0360)
Estimated differences in slopes between 1860--99 and 1820-59			
Charles Darwin	0.021***	Darwin	0.307***
	(0.004)		(0.047)
Alfred Wallace	0.003***	Wallace	0.104***
	(0.0003)		(0.037)
Robert Chambers	-0.001	Chambers	-0.147***
	(0.001)		(0.042)
Jean Baptiste Lamarck	0.000	Lamarck	0.036**
	(0.000)		(0.014)
Observations	320	Observations	320
R-squared	0.707	R-squared	0.921

Notes: The table displays coefficient estimates from spline regressions of the number of occurrences (per million words) of the names Charles Darwin (omitted category), Alfred Russel Wallace, Robert Chambers and Jean Baptiste Lamarck (Column (1)); and Darwin (omitted category), Wallace, Chambers and Lamarck (Column (2)). The coefficient estimates on the binary indicators for each name are not reported in the table. When we consider both the first and last names (Column (1)), we include different combinations of the full names of the four scientists: Alfred Russel Wallace, Alfred Wallace, Charles Darwin, Charles Robert Darwin, Robert Chambers, Jean-Baptiste Lamarck, Jean-Baptiste de Lamarck, Jean Baptiste Lamarck, and Jean Baptiste de Lamarck. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Spline Regression Analyses (eight knots) – Frequency of Darwinian Concepts in US Congressional Records and UK Parliamentary Debates

Panel A: UK Parliament

Word/phrase:	Natural Selection	Evolution	Selection	Competition	Survival	Adaptation
	(1)	(2)	(3)	(4)	(5)	(6)
Slope(1820-29)	0.001	0.042	-2.806	-10.826*	-0.000	0.165
	(0.000)	(0.059)	(2.317)	(5.498)	(0.000)	(0.237)
Slope(1830-39)	-0.002	-0.062	4.730**	7.669	0.000	-0.052
	(0.001)	(0.068)	(2.107)	(5.777)	(0.000)	(0.208)
Slope(1840-49)	0.007*	0.119	-2.391	10.579*	-0.001	0.218
	(0.004)	(0.077)	(1.974)	(5.953)	(0.001)	(0.155)
Slope(1850-59)	-0.026*	-0.078	2.703	-10.189*	0.005	0.031
	(0.014)	(0.071)	(2.169)	(5.801)	(0.005)	(0.161)
Slope(1860-69)	0.141***	0.057	5.164	-2.302	-0.019	0.083
	(0.045)	(0.082)	(5.309)	(3.955)	(0.017)	(0.194)
Slope(1870-79)	-0.017	0.167	-4.406	-0.104	0.128**	-0.408**
	(0.096)	(0.184)	(4.553)	(3.578)	(0.063)	(0.158)
Slope(1880-89)	-0.066	-0.218	-2.289	-7.569	0.119	0.023
	(0.081)	(0.216)	(2.018)	(5.033)	(0.138)	(0.135)
Slope(1890-99)	0.013	0.189	1.234	17.535**	0.015	0.163
	(0.038)	(0.126)	(2.039)	(6.719)	(0.146)	(0.117)
Observations	80	80	80	80	80	80
R-squared	0.271	0.129	0.253	0.315	0.536	0.223

Notes: For each word and phrase, the estimates refer to the slope of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with eight knots at 19, 29, …, 89. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Panel B: US Congress

Word/phrase:	Natural Selection	Evolution	Selection	Competition	Survival	Adaptation
	(1)	(2)	(3)	(4)	(5)	(6)
Slope(1820-29)	-0.000	0.169	2.908	-1.334	-0.002	0.548
	(0.000)	(0.105)	(2.146)	(8.500)	(0.002)	(0.352)
Slope(1830-39)	0.000	-0.111	-3.557***	-12.213*	0.006	-0.501*
	(0.000)	(0.077)	(1.130)	(6.148)	(0.005)	(0.281)
Slope(1840-49)	-0.001	0.065	0.326	1.669	-0.002	0.412**
	(0.001)	(0.045)	(1.035)	(3.567)	(0.003)	(0.199)
Slope(1850-59)	0.003	0.007	-0.760	-4.716	0.003	-0.302*
	(0.003)	(0.065)	(0.722)	(3.451)	(0.007)	(0.177)
Slope(1860-69)	0.003	-0.066	1.450*	3.532	-0.030*	-0.177*
	(0.004)	(0.067)	(0.797)	(2.500)	(0.016)	(0.098)
Slope(1870-79)	-0.000	0.011	-1.458*	0.764	0.133***	0.006
	(0.010)	(0.059)	(0.753)	(3.671)	(0.048)	(0.067)
Slope(1880-89)	0.035**	0.358**	1.602*	19.141***	0.140	0.146
	(0.013)	(0.169)	(0.937)	(5.115)	(0.091)	(0.096)
Slope(1890-99)	-0.019	1.067***	-0.606	-15.709**	0.061	-0.126
	(0.020)	(0.255)	(1.117)	(6.805)	(0.133)	(0.114)
Observations	80	80	80	80	80	80
R-squared	0.447	0.725	0.271	0.393	0.674	0.285

Notes: For each word and phrase, the estimates refer to the slope of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with eight knots at 19, 29, …, 89. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Spline Regression Analyses (one knot) – Frequency of Darwinian Concepts. Newey-West correction.

Word/phrase:	Natural Selection	Evolution	Selection	Competition	Survival	Adaptation
	(1)	(2)	(3)	(4)	(5)	(6)
Slope(1820-59)	0.027**	-0.003	0.177***	-0.071***	-0.020***	0.117***
	(0.012)	(0.027)	(0.023)	(0.019)	(0.006)	(0.017)
Slope(1860-99)	0.165***	0.950***	0.372***	0.269***	0.171***	0.002
	(0.022)	(0.054)	(0.038)	(0.021)	(0.008)	(0.010)
Observations	80	80	80	80	80	80

Notes: For each word and phrase, the two estimates Slope(1820-59) and Slope(1860-99) refer to the slopes of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with one knot at 59. The regression equation for a given word w is: $y_{wt} = \alpha_w + \beta_{1w} \min(t, 59) + \beta_{1w} (\max(t, 59) - 59) + \varepsilon_{wt}$. Slope(1860-99)-Slope(1820-59) represents the estimate of the difference between the two slopes. Newey-West standard errors (with one autocorrelation lag with respect to time) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Spline Regression Analyses (one knot) – Frequency of Darwinian Concepts, Separate for Fiction and Non-fiction Books. Newey-West correction.

Word/phrase:	Natural Selection		Evolution		Selection	
	Non-fiction	Fiction	Non-fiction	Fiction	Non-fiction	Fiction
	(1)	(2)	(3)	(4)	(5)	(6)
Slope(1820-59)	0.027**	0.005***	-0.009	-0.022**	0.183***	0.031
	(0.012)	(0.002)	(0.030)	(0.010)	(0.024)	(0.025)
Slope(1860-99)	0.167***	0.012***	1.028***	0.173***	0.412***	0.040**
	(0.022)	(0.003)	(0.060)	(0.023)	(0.041)	(0.018)
Observations	80	80	80	80	80	80

Word/phrase:	Competition		Survival		Adaptation	
	Non-fiction	Fiction	Non-fiction	Fiction	Non-fiction	Fiction
	(7)	(8)	(9)	(10)	(11)	(12)
1820-59	-0.074***	-0.008	-0.021***	-0.013***	0.122***	0.025***
	(0.020)	(0.021)	(0.006)	(0.005)	(0.017)	(0.009)
1860-99	0.300***	-0.000	0.178***	0.096***	0.010	-0.012*
	(0.023)	(0.015)	(0.009)	(0.010)	(0.011)	(0.007)
Observations	80	80	80	80	80	80

Notes: For each word and phrase, the two estimates refer to the slope of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with one knot at 59. The regression equation for a given word *w* is: $y_{wt} = \alpha_w + \beta_{1w} \min(t, 59) + \beta_{1w} (\max(t, 59) - 59) + \varepsilon_{wt}$. Slope(1860-99)-Slope(1820-59) represents the estimate of the difference between the two slopes. Newey-West standard errors (with one autocorrelation lag with respect to time) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Spline Regression Analyses (eight knots) – Average Frequency of Darwinian and Generic Words. Newey-West correction.

Word/phrase and sample:	Darwinian words: Non-fiction	Darwinian words: Fiction	Darwinian words: Full sample	Generic words: Full Sample
	(1)	(2)	(3)	(4)
Slope(1820-29)	0.142***	0.104**	0.142***	0.990***
	(0.038)	(0.046)	(0.037)	(0.281)
Slope(1830-39)	0.073**	0.025	0.067**	0.333
	(0.028)	(0.065)	(0.028)	(0.248)
Slope(1840-49)	0.065**	-0.049	0.060**	-0.185
	(0.030)	(0.066)	(0.030)	(0.206)
Slope(1850-59)	-0.031	0.044	-0.028	-0.042
	(0.039)	(0.027)	(0.037)	(0.182)
Slope(1860-69)	0.339***	-0.058**	0.312***	0.138
	(0.064)	(0.023)	(0.061)	(0.174)
Slope(1870-79)	0.213***	0.150***	0.219***	0.088
	(0.069)	(0.035)	(0.065)	(0.137)
Slope(1880-89)	0.565***	0.081*	0.512***	0.275**
	(0.061)	(0.041)	(0.057)	(0.116)
Slope(1890-99)	0.336*	-0.047	0.273	-0.313
	(0.176)	(0.054)	(0.164)	(0.194)
Observations	80	80	80	80

Notes: Each yearly observation is the average frequency (per million words) of the six main Darwinian words and phrases (full sample as well as separate between fiction and non-fiction) and of the 99 most frequent nouns in *On the Origin of Species* that we use as our control group of generic words. The estimates refer to the slope of the best linear fit from a spline regression of frequency (per million words) on years from 1820 to 1899, expressed as t=20, 21, …, 99, with eight knots at 19, 29, …, 89.

The regression equation for a given word w is: $y_{wt} = \alpha_w + \beta_{1w} \min\{t, 29\} + \beta_{2w}(Max\{\min\{t, 39\}, 29\} - 29) + \beta_{3w}(Max\{\min\{t, 49\}, 39\} - 39) + \beta_{4w}(Max\{\min\{t, 59\}, 49\} - 49) + \beta_{5w}(Max\{\min\{t, 69\}, 59\} - 99) + \beta_{6w}(Max\{\min\{t, 79\}, 69\} - 69) + \beta_{7w}(Max\{\min\{t, 89\}, 79\} - 79) + \beta_{8w}(Max\{\min\{t, 99\}, 89\} - 89) + \varepsilon_{wt}$.

Newey-West standard errors (with one autocorrelation lag with respect to time) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.