Supplementary Materials—Order Matters: Alphabetizing In-Text Citations Biases Citation Rates

Jeffrey R. Stevens and Juan F. Duque

Supplementary Analyses

Article-Level Analyses

For the article-level analyses, we combined data sets 1 and 2 and removed all articles with zero citations. Because our citation count data were highly skewed with a heavy tail, we log-transformed the data to improve normality and applied a Linear Mixed Model (LMM), using the *lme4* package in R. We included both journal and year (rescaled from 2000-2015 to 0-15) as random effects and converted letters to numbers (0-25). We tested models with and without random slopes for journal and year, but including random slopes did not improve model fit ($\chi^2 < 4.4$, BF<0.0001) and hindered the optimization of LMM parameter estimates. Thus, we report results from models without the random slopes.

We conducted two LMMs with letter and field as fixed effects: one model included their interaction and the other included only the main effects. We then extracted the BIC values for both models. Using the conversion $\text{BF} = e^{(BIC_{nointeraction} - BIC_{interaction})/2}$, we calculated the Bayes factor for the model with the interaction compared to the model without the interaction.

Our proposed mechanism for how the alphabetical citation bias arises depends on articles already being cited. Given that, as articles are cited more and more, the likelihood of being included in a list of in-text citations increases, our mechanism predicts that the bias should be stronger among more highly cited articles. To explore the degree to which the high citation count articles drive the alphabetical citation bias, we calculated the LMM analysis with different subsets of the data. First, we created 10 subsets of the data based on the top percentile. Percentile 0 included all data, percentile 10 excluded the bottom 10% and included the top 90% of articles with the most citations, and so on. We selectively removed the bottom percentiles because we expected larger alphabetical citation biases for articles with higher citation counts. Moreover, most articles are either not cited or have very few citation counts. For example, in our pooled dataset, excluding 50% of the least cited articles (50th percentile) resulted in articles that had 10-14 (depending on the letter) or more citation counts. Thus, these articles are not outliers but the predicted subset of the data in which the bias should occur. For each percentile, we calculated LMMs with (1) the main effects and interaction for letter (A-Z) and field (psychology, biology, and geosciences) and (2) just the main effects. We then calculated Bayes factors for the model with the interaction compared to the model without. Thus, this Bayes factor is the strength of evidence for or against the presence of an interaction. Figure S6 shows the Bayes factor for each percentile. As the data are restricted to more highly cited articles, Bayes factors for the interaction between letter and field increase. Examination of the estimated marginal means of the linear trends showed the alphabetical citation bias in psychology but not biology and geoscience. Evidence for this interaction begins at percentile 50, which includes articles with at least 10-14 citations.

R Packages

This project used R (Version 3.5.1; R Core Team, 2018) and the R-packages *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2018), *car* (Version 3.0.2; Fox & Weisberg, 2011), *foreach* (Version 1.4.4; Microsoft & Weston, 2017), *lme4* (Version 1.1.18.1; Bates, Mächler, Bolker, & Walker, 2015), *papaja* (Version 0.1.0.9842; Aust & Barth, 2017), *scales* (Version 1.0.0; Wickham, 2017), and *tidyverse* (Version 1.2.1; Wickham, 2017).

Table S1: Journal information for data set 1

Journal	Field	Citation style[*]	Articles
American Psychologist	Psychology	Alphabetical	2,080
Animal Behaviour[a]	Biology	Alphabetical	5,073
Animal Cognition	Psychology	Alphabetical	1,008
Annual Review of Ecology, Evolution, and Systematics[b]	Biology	Numerical	403
Annual Review of Psychology	Psychology	Alphabetical	404
Behavioral Ecology[c]	Biology	Chronological	2,540
Behavioral Ecology and Sociobiology	Biology	Chronological	2,730
Behaviour	Biology	Chronological	1,348
Behavioural Processes[d]	Both	Chronological	0
Biological Reviews	Biology	Chronological	565
Current Directions in Psychological Science	Psychology	Alphabetical	1,066
Ethology	Biology	Chronological	1,735
Journal of Comparative Psychology	Psychology	Alphabetical	774
Journal of Ethology	Biology	Chronological	631
Journal of Experimental Psychology: Animal Learning and Cognition[e]	Psychology	Alphabetical	636
Journal of the Experimental Analysis of Behavior	Psychology	Alphabetical	841
Learning and Behavior	Psychology	Alphabetical	478
Perspectives on Psychological Science[f]	Psychology	Alphabetical	594
Philosophical Transactions of the Royal Society[g]	Biology	Numerical	4,520
PLOS Biology[h]	Biology	Numerical	3,054
Proceedings of the Royal Society[g]	Biology	Numerical	7,469
Psychological Bulletin	Psychology	Alphabetical	804
Psychological Review	Psychology	Alphabetical	736
Psychological Science	Psychology	Alphabetical	3,094
Quarterly Review of Biology	Biology	Chronological	303
Trends in Cognitive Sciences	Psychology	Numerical	1,907
Trends in Ecology and Evolution	Biology	Numerical	2,135

[*]Citation style as of 2015. [a]Switched from chronological to alphabetical in Jan 2014. [b]Switched from numerical to alphabetical in 2001 and changed its name in 2003—formerly Annual Reviews of Ecology and Systematics. [c]Switched from alphabetical to chronological in Jul 2006. [d]Not included in analysis since covers both biology and psychology fields. [e]Changed its name in 2014—formerly Journal of Experimental Psychology: Animal Behavior Processes. [f]Started in 2006. [g]Switched from chronological to numerical in Jan 2011. [h]Started in 2003 and switched from chronological to numerical in Jan 2005.

Table S2: Journal information for data set 2

Journal	Field	Citation style*	Articles
American Journal of Orthopsychiatry	Psychology	Alphabetical	977
Asian American Journal of Psychology	Psychology	Alphabetical	200
Atmospheric Science Letters	Geoscience	Chronological	556
Attention, Perception, & Psychophysics	Psychology	Alphabetical	1,285
Behavioral Neuroscience	Psychology	Alphabetical	1,952
Behavior Research Methods	Psychology	Alphabetical	1,199
Bulletin of the Seismological Society of America	Geoscience	Chronological	3,505
Canadian Journal of Experimental Psychology	Psychology	Alphabetical	2,840
Canadian Psychology	Psychology	Alphabetical	809
Cognitive, Affective, & Behavioral Neuroscience	Psychology	Alphabetical	705
Contemporary Psychology: APA Review of Books	Psychology	Alphabetical	1,397
Cultural Diversity & Ethnic Minority Psychology	Psychology	Alphabetical	475
Developmental Psychology	Psychology	Alphabetical	1,578
Dreaming	Psychology	Alphabetical	236
Earth Interactions	Geoscience	Chronological	279
Emotion	Psychology	Alphabetical	1,249
Environmental and Engineering Geoscience	Geoscience	Chronological	428
Experimental and Clinical Psychopharmacology	Psychology	Alphabetical	837
Geochemistry: Exploration, Environment, Analysis	Geoscience	Chronological	353
Geological Society of America Bulletin	Geoscience	Chronological	1,808
Geology	Geoscience	Chronological	4,756
Geosphere	Geoscience	Chronological	657
Group Dynamics: Theory, Research, and Practice	Psychology	Alphabetical	340
Health Psychology	Psychology	Alphabetical	1,663
History of Psychology	Psychology	Alphabetical	201
International Journal of Climatology	Geoscience	Chronological	2,778
Italian Journal of Geosciences	Geoscience	Chronological	209
Journal of Abnormal Psychology	Psychology	Alphabetical	1,340
Journal of Applied Meteorology and Climatology	Geoscience	Chronological	1,678
Journal of Applied Psychology	Psychology	Alphabetical	1,620
Journal of Atmospheric and Oceanic Technology	Geoscience	Chronological	2,537
Journal of Climate	Geoscience	Chronological	6,527
Journal of Consulting and Clinical Psychology	Psychology	Alphabetical	1,747
Journal of Counseling Psychology	Psychology	Alphabetical	836
Journal of Diversity in Higher Education	Psychology	Alphabetical	157
Journal of Educational Psychology	Psychology	Alphabetical	1,100
Journal of Experimental Psychology: Applied	Psychology	Alphabetical	423
Journal of Experimental Psychology: General	Psychology	Alphabetical	894
Journal of Experimental Psychology: Human Perception and Performance	Psychology	Alphabetical	1,895
Journal of Experimental Psychology: Learning, Memory, and Cognition	Psychology	Alphabetical	1,887
Journal of Family Psychology	Psychology	Alphabetical	1,296
Journal of Hydrometeorology	Geoscience	Chronological	1,500
Journal of Micropalaeontology	Geoscience	Chronological	283
Journal of Occupational Health Psychology	Psychology	Alphabetical	409
Journal of Personality and Social Psychology	Psychology	Alphabetical	2,299
Journal of Physical Oceanography	Geoscience	Chronological	2,873
Journal of the Geological Society	Geoscience	Chronological	1,452
Law and Human Behavior	Psychology	Alphabetical	691
Lithosphere	Geoscience	Chronological	264
Memory & Cognition	Psychology	Alphabetical	1,935
Meteorological Applications	Geoscience	Chronological	770
Military Psychology	Psychology	Alphabetical	464
Monthly Weather Review	Geoscience	Chronological	3,287
Neuropsychology	Psychology	Alphabetical	1,216

|			
Personality Disorders: Theory, Research, and Treatment	Psychology	Alphabetical	261
Petroleum Geoscience	Geoscience	Chronological	502
Proceedings of the Yorkshire Geological Society	Geoscience	Chronological	199
Professional Psychology: Research and Practice	Psychology	Alphabetical	1,371
Psychiatric Rehabilitation Journal	Psychology	Alphabetical	1,041
Psychoanalytic Psychology	Psychology	Alphabetical	807
Psychological Assessment	Psychology	Alphabetical	1,203
Psychological Methods	Psychology	Alphabetical	481
Psychological Services	Psychology	Alphabetical	289
Psychological Trauma: Theory, Research, Practice, and Policy	Psychology	Alphabetical	425
Psychology and Aging	Psychology	Alphabetical	1,323
Psychology of Addictive Behaviors	Psychology	Alphabetical	1,264
Psychology of Aesthetics, Creativity, and the Arts	Psychology	Alphabetical	334
Psychology of Men & Masculinity	Psychology	Alphabetical	309
Psychology of Religion and Spirituality	Psychology	Alphabetical	198
Psychology of Violence	Psychology	Alphabetical	178
Psychology, Public Policy, and Law	Psychology	Alphabetical	398
Psychotherapy	Psychology	Alphabetical	919
Quarterly Journal of Engineering Geology and Hydrogeology	Geoscience	Chronological	581
Quarterly Journal of the Royal Meteorological Society	Geoscience	Chronological	2,769
Rehabilitation Psychology	Psychology	Alphabetical	778
Review of General Psychology	Psychology	Alphabetical	407
School Psychology Quarterly	Psychology	Alphabetical	482
Scottish Journal of Geology	Geoscience	Chronological	269
Seismological Research Letters	Geoscience	Chronological	1,065
South African Journal of Geology	Geoscience	Chronological	498
Training and Education in Professional Psychology	Psychology	Alphabetical	325
Vadose Zone Journal	Geoscience	Chronological	1,498
Weather and Forecasting	Geoscience	Chronological	1,373
Weather, Climate, and Society	Geoscience	Chronological	200

*Citation style as of 2015.
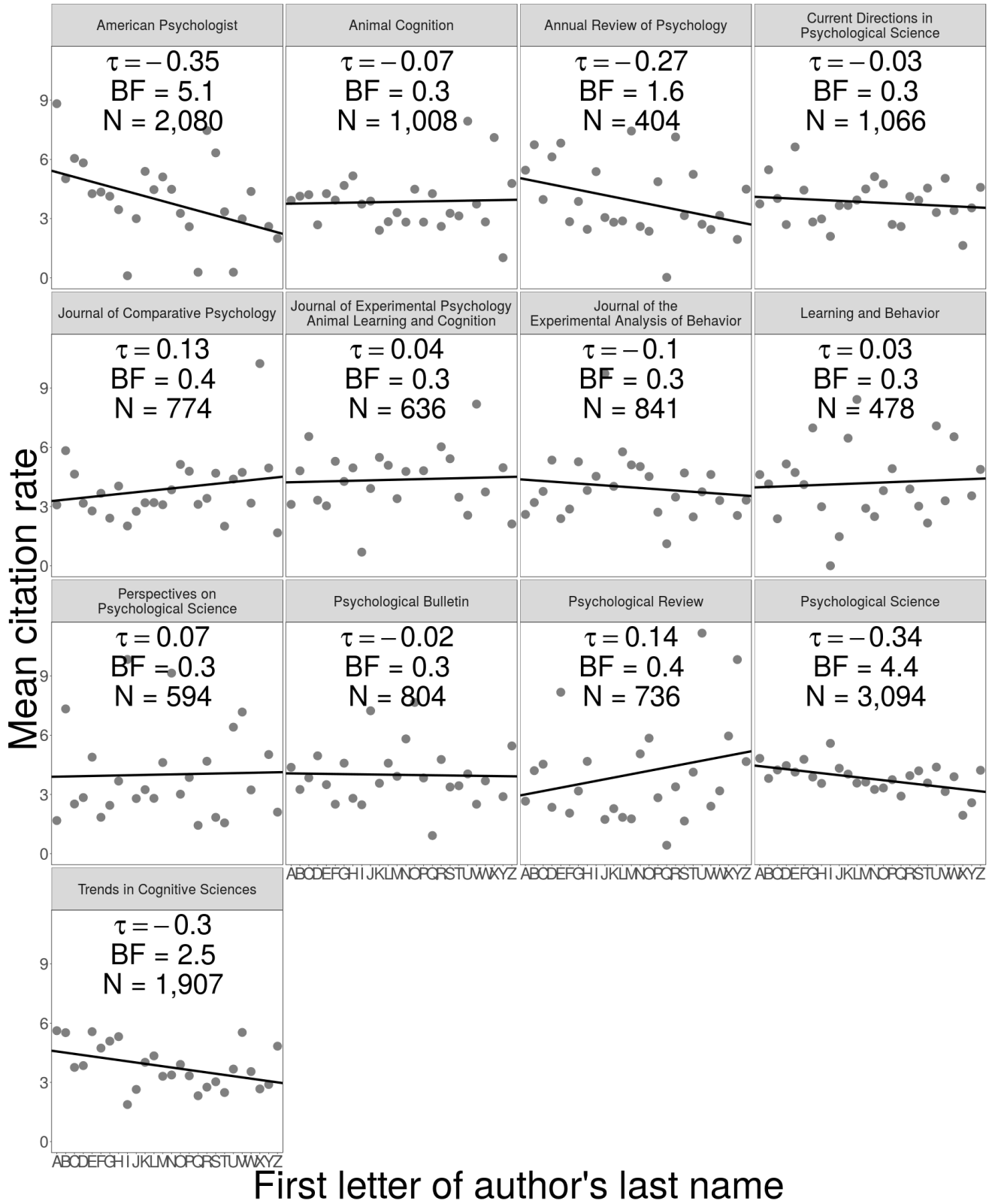
Figure S1: Citation rate per psychology journal for data set 1. Citation rate refers to the mean percentage of total citations for each letter. τ denotes the non-parametric Kendall's τ, but the line slope is determined by a parametric linear regression.
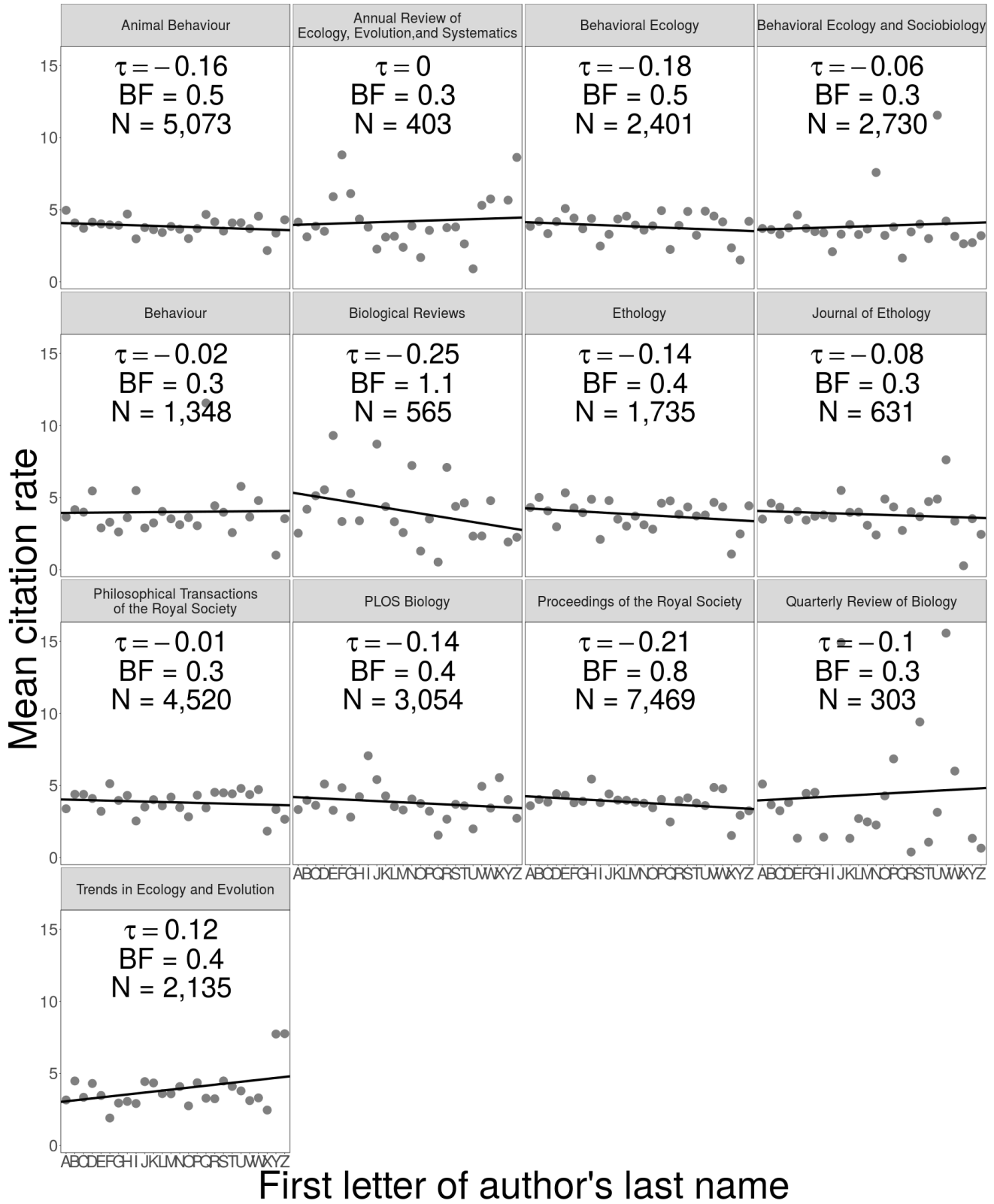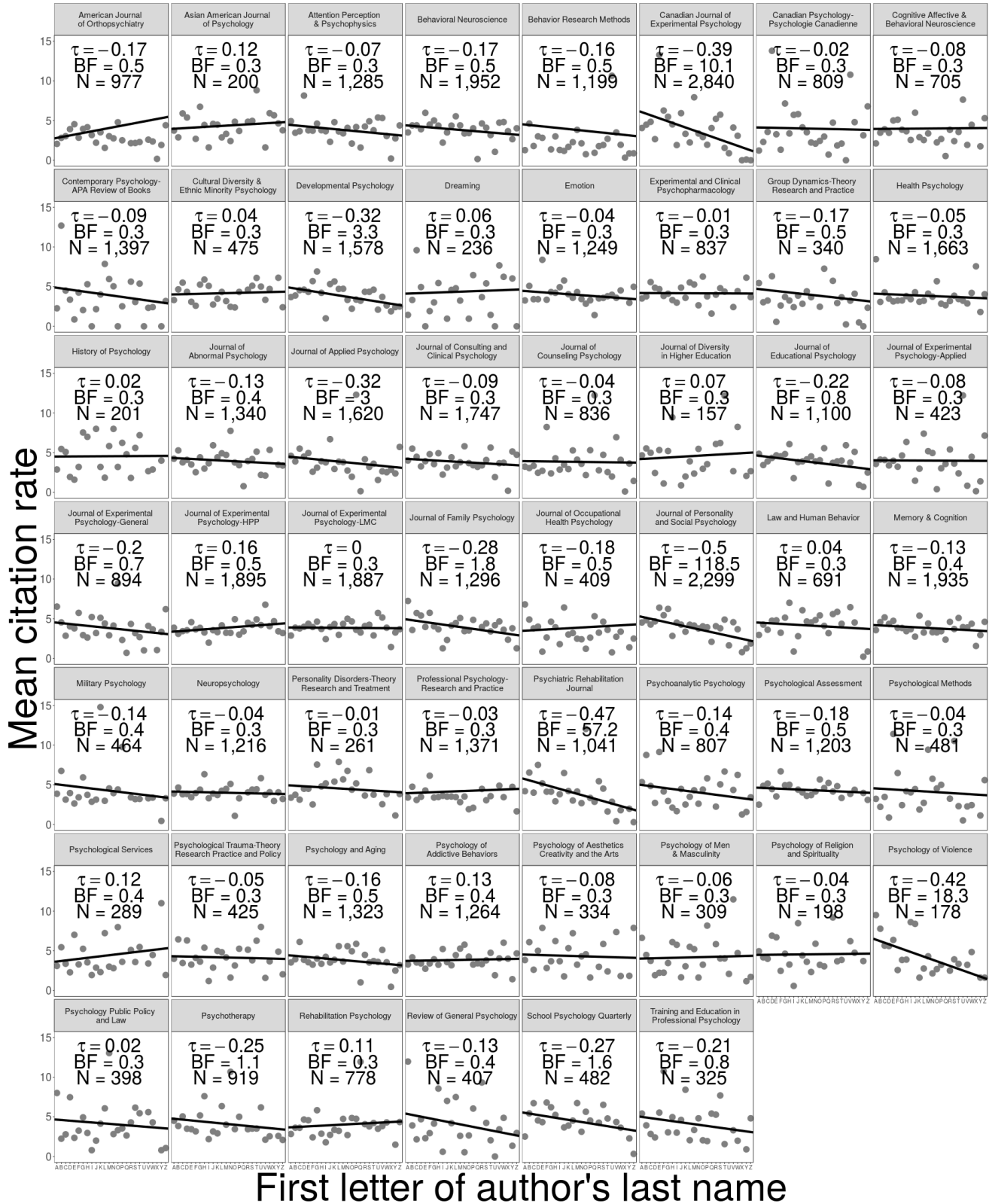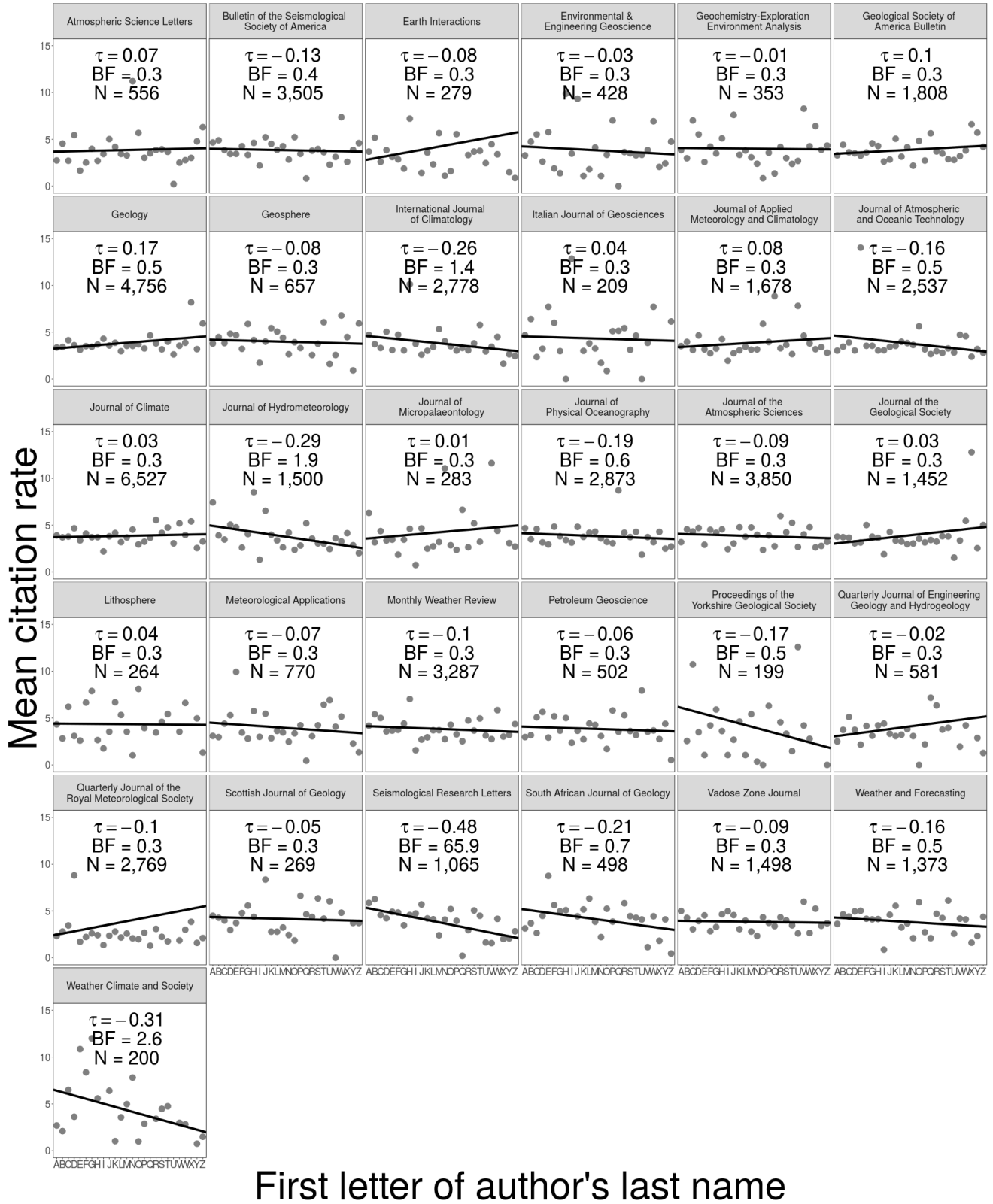
Figure S2: Citation rate per biology journa for data set 1l. Citation rate refers to the mean percentage of total citations for each letter. τ denotes the non-parametric Kendall's τ, but the line slope is determined by a parametric linear regression.

Figure S3: Citation rate per psychology journal for data set 2. Citation rate refers to the mean percentage of total citations for each letter. τ denotes the non-parametric Kendall's τ, but the line slope is determined by a parametric linear regression. Some data points are not shown because they extend beyond the illustrated y-axis range.
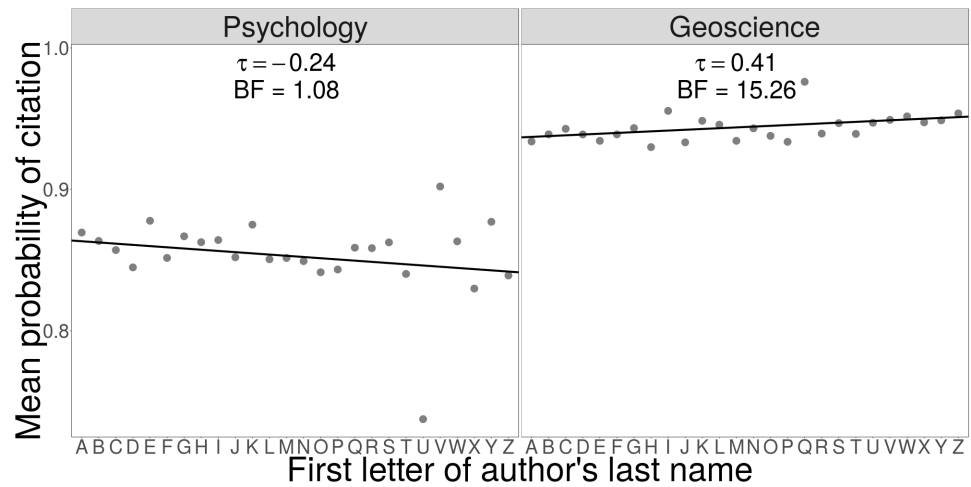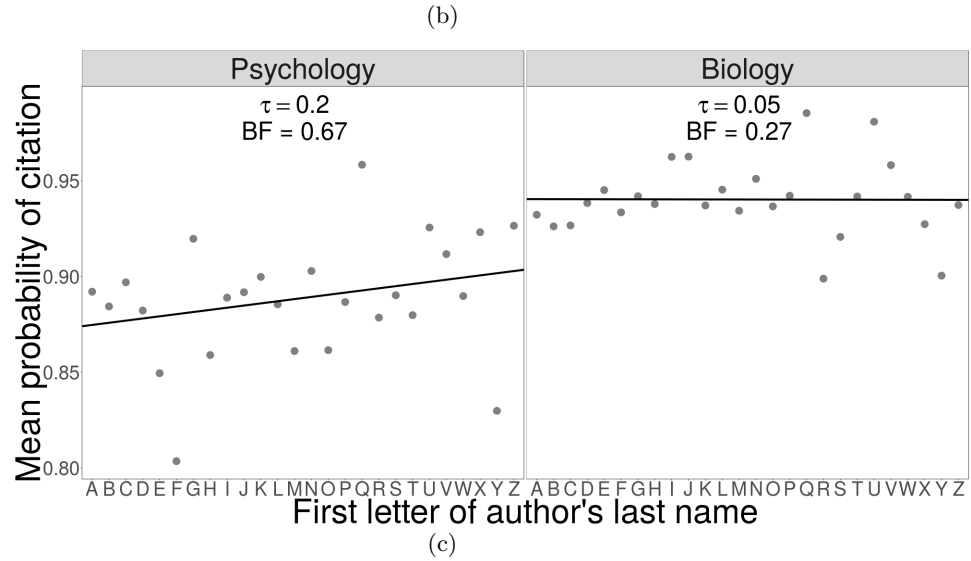
Figure S4: Citation rate per geoscience journal for data set 2. Citation rate refers to the mean percentage of total citations for each letter. τ denotes the non-parametric Kendall's τ, but the line slope is determined by a parametric linear regression. Some data points are not shown because they extend beyond the illustrated y-axis range.
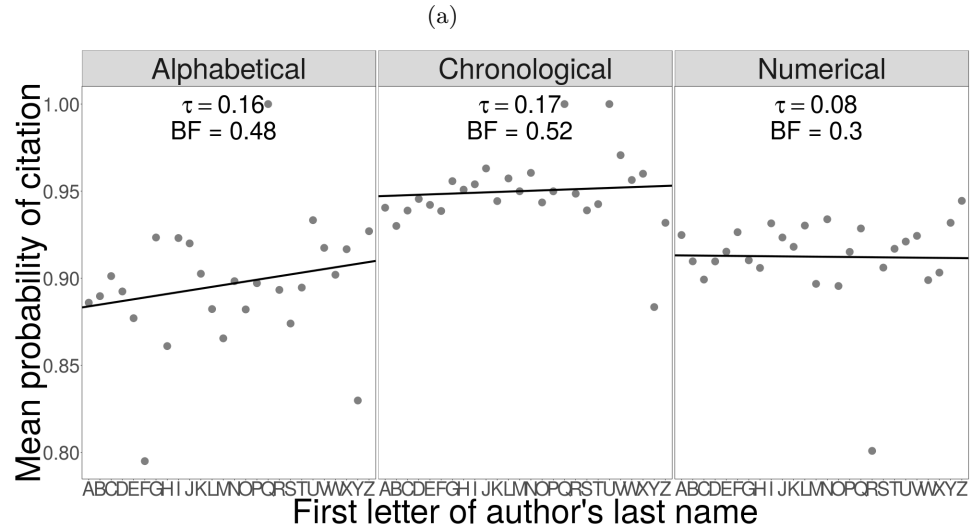
Figure S5: Probability of citation per citation style or field. Probability of citation refers to the mean proportion of articles that have been cited at least once for each letter grouped by (a) citation style in data set 1, (b) field in data set 1, and (c) field in data set 2.
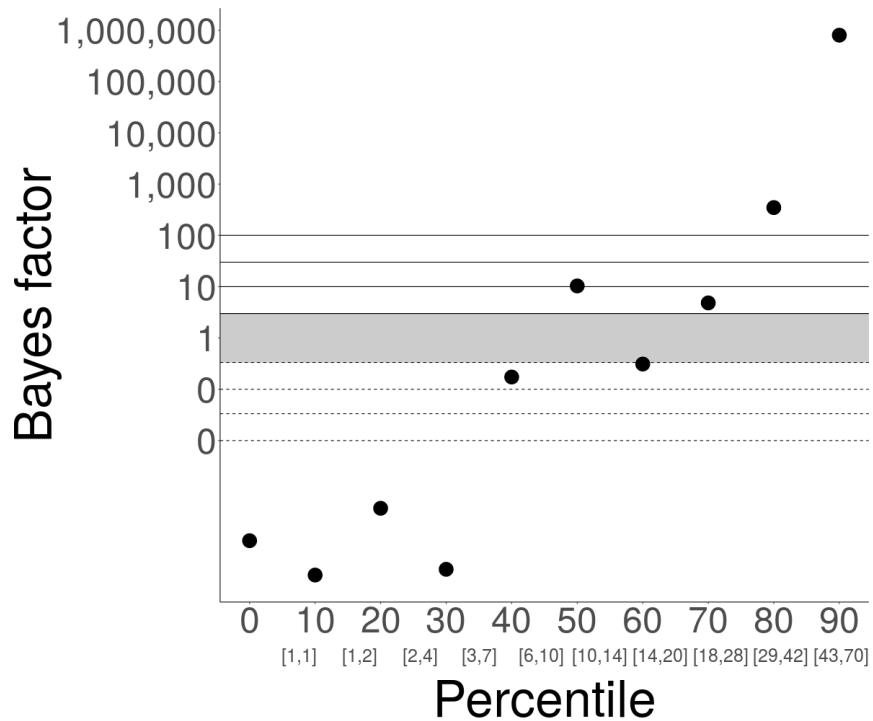
9

Figure S6: Bayes factors for alphabetical citation bias across fields. We extracted the top percentiles (0-90th) of data based on number of citations (0th = all data, 90th = top 10 percent of most cited articles per letter). Numbers in brackets below percentiles give the range of the minimum number of citations for that percentile across letters. For example, at the 50th percentile, the analysis includes the articles with 10-14 or more citations, depending on the letter. In general, Bayes factors for the presence of an interaction between letter and field on citation count increased with higher citation count percentiles of data. Horizontal lines represent Bayes factor thresholds 3, 10, 30, and 100 (solid lines) and 1/3, 1/10, 1/30, and 1/100 (dashed lines). Shaded area denotes no evidence for or against an interaction. Thus, data points above the shaded area indicate an interaction, and data points below the area indicate no interaction.

AsPredicted
Pre-Registration made easy

Order Matters: The Psychology of Citations---Data set 2: APA vs. geology (#3152)

Author(s)
Jeffrey Stevens (University of Nebraska-Lincoln) - jstevens5@unl.edu
Juan Duque (University of Nebraska-Lincoln) - jfduque89@gmail.com

1) What's the main question being asked or hypothesis being tested in this study?
Does alphabetical ordering of in-text citations cause a bias toward higher citation counts for articles whose first author's surname has a letter earlier in the alphabet. We hypothesize that articles with alphabetically ordered citations will show a negative relationship between citation counts and position in the alphabet of the letter of the first author's last name. We also hypothesize that chronologically ordered citations will show no relationship with author letter and, therefore, will have different slopes from alphabetically ordered citations.

2) Describe the key dependent variable(s) specifying how they will be measured.
Mean number of times that articles are cited per letter of first author's surname, corrected for the total number of articles with first author surnames with that letter and converted into a percentage.

3) How many and which conditions will participants be assigned to?
No participants are used in this analysis, just information about article author names and the citation style for the journal that published the article. We will compare two citation styles: alphabetically ordered in-text citations and chronologically ordered in-text citations. We will use psychology (APA) journals for alphabetical citations and geology/meteorology journals for chronological citations.

4) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
For all data, then for each citation type
  Aggregate the total number of citations per letter
  Aggregate the total number of articles per letter
  Calculate number of citations per article per letter
  Convert to percentage (percent citations per article per letter–citation index hereafter)
  Check for outlier letters
    Calculate modified Z score (Iglewicz & Hoaglin, 1993) for citation index of each letter (only for overall data, then apply to each citation type)
    Remove letter if Z score > 3.5
Calculate Bayesian linear regression of citation index ~ letter separately for each citation type to get Bayes factor for slope vs. intercept comparison
Calculate Bayesian linear regression of citation index ~ letter + citation type for null model
Calculate Bayesian linear regression of citation index ~ letter * citation type for alternative model
Divide Bayes factors to determine Bayes factor of interaction, which indicates slope differences
For Bayesian linear regressions, we will use the default hierarchical prior distribution from lmBF that is weakly informative (Rouder & Morey, 2012).

5) Any secondary analyses?
We will also use prior distributions based on an existing data set.
 An alternative way to test for slope differences:
Calculate frequentist linear regression of citation index ~ letter separately for each citation type to get slope means and standard errors
Calculate z-test to compare slopes of two citation types (Paternoster et al. 1998)
Convert z statistic to Bayesian t-test to test for slope differences

6) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect at least 50,000 articles of each citation style type. If we fail to find a Bayes factor for the comparison of the alphabetical citation style slope to 0 or for the comparison of the slope differences to be greater than 10 or less than 0.10, we will add more articles to both citation style types until this threshold is met. If the threshold is not met at 75,000 articles, we will stop and report the Bayes factors.

7) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)
We may collect 50,000 articles from neuroscience with chronological citations, but the effect may not be as strong with them because of the carry-over effects with psychologists publishing in neuroscience journals.

8) Have any data been collected for this study already?
No, no data have been collected for this study yet

References

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown.* https://github.com/crsh/papaja.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (Second ed.). Thousand Oaks CA: Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Microsoft, & Weston, S. (2017). *foreach: Provides foreach looping construct for R.* https://CRAN.R-project.org/package=foreach.

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs.* https://CRAN.R-project.org/package=BayesFactor.

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, H. (2017). *scales: Scale functions for visualization.* https://CRAN.R-project.org/package=scales.

Wickham, H. (2017). *tidyverse: Easily install and load the "tidyverse".* https://CRAN.R-project.org/package=tidyverse.