SUPPLEMENTARY INFORMATION

For data, code, figures, visualizations, and other files, go to:

https://osf.io/qf8zv/?view_only=529f6c8186664f18a0492e95c3893b57

Complement to Main Text

Methods

Web Scraping

We used Python 3.5.2 BeautifulSoup module to scrape the blog data. We then used NLTK module to clean and preprocess it. In particular, we: 1) combined hyphenated words into a single word, tokenized the data (i.e., divided it into words), changed all words to lower case, lemmatized words (i.e., remove inflectional endings to retain only the base form of a word; e.g., changing “be” to “is” or “walking” to “walk”), deleted stopwords (i.e., common words that provide little information; e.g., “the”), and removed punctuation and numbers. Only posts in English were included, as determined by an algorithm that checked the number of English words in the document. Other variables scraped where available included post tags, which were not analyzed in depth, but top tags are presented in a separate document in the repository. Blog citations might differ slightly from the sum of post citations since some citations were not attached to specific posts. We analyze blog citations for univariate descriptive statistics, and post citations for relationships. Person names were obtained in Python’s NLTK 3.2.5 module using the Stanford Named Entity Recognizer 3.8.0, 3 class model (Finkel, Grenager, & Manning, 2005). We used SharedCount (https://www.sharedcount.com/) to obtain social media indicators.

Other analyses’ cleaning used slight variations of the workflow presented in the main text. The list of comment authors retained only alphanumeric and space characters, and words were not changed to lower case. Length variables included all words (no stopwords deleted), except when using as offsets for topic or sentiment analyses, which used preprocessed post/comment length.

For analyses on the *Behavioral and Brain Sciences* articles we obtained all articles from their online versions. Then, we applied the same topic model trained on the blog posts to the articles' text (see online repository for all data and a topic breakdown). In addition, we used an automated gender classifier on the authors' first names (https://genderize.io/). The algorithm is trained on social media profiles and it classified 95% of the names (see online repository for all classification data and code).

Sentiment scores were based on a dictionary approach, using the NRC Emotion Lexicon (Mohammad & Turney, 2013). The NRC Emotion Lexicon is a list of words and their associations with positive and negative sentiment. The association scores are binary (1 indicates association, 0 indicates no association), which were manually annotated by Amazon's Mechanical Turkers. To construct post/comment emotion scores, we counted word overlaps between post/comment contents and the dictionary using Python. Thus, emotion scores were number of positive (e.g., "ability", "important", "respectful") and negative (e.g., "abuse", "lax", "polemic") words per post/comment, and were analyzed with mixed negative binomial regressions. See http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm for access to the full lexicon.

The sentiment score procedure utilized here has several limitations. The dictionary-based approach was unable to capture patterns proper of informal communication (e.g., sarcasm), negations (e.g., "She is not smart" would be rated as positive because words are evaluated individually), or other predictive features at the sentence-level. Additionally, although some of the discussion surrounding online communication has revolved around tone, the current sentiment analysis might be a poor indicator of the frequency and impact of civil or uncivil discussion. In other words, the sentiment analysis simply indicates the relative number of words rated as positive or negative when evaluated in isolation by human coders. Future attempts to study tone should focus on coded methods (Jamieson, 1999), domain-specific natural language processing models, dependency parsing (Jurafsky & Martin, 2016), and qualitative analyses.

Date analyses

All models with date as a predictor included in the Supplement had post-level outcomes, and thus we used the same approach as for other predictors, with date standardized at the group level (i.e., how recent is the post compared to other posts within the blog?) and at the grand mean (i.e., how recent is the average post for the blog?; this is related but not identical to the age of the blog). In these, the date predictor was the R default, coded as the number of days since 1970-01-01.

LDA Topics

For an introduction to LDA from which we based our simplified description of the algorithm, see Griffiths, Steyvers, and Tenenbaum (2007), or https://eight2late.wordpress.com/2015/09/29/a-gentle-introduction-to-topic-modeling-using-r/. As a very simplified summary, the algorithm works by initially randomly assigning each word in each document to a topic, creating preliminary distributions of words over topics and topics over documents. Then, the algorithm iterates multiple times over the different documents and reassigns each word into a topic based on how likely the word is to belong to that new topic (based on the current distribution of words over topics) and how likely the topic is to occur in that document (based on the current distribution of topics over documents). Thus, the model iteratively updates a document words' topic assignment assuming that a word occurring often in a topic is more likely to belong to that topic, and that a document with many words from a topic is more likely to have more words from that topic.

The LDA model was run on the preprocessed data. In addition, for LDA we used only words that appeared in at least 2 posts, thus removing possible typos, nonwords, etc. Choice of an appropriate number of topics was based on the ldatuning R package, attempting to identify a variety of topic number estimates (from 2 to 45) and choosing from those that performed well on the different metrics provided by the package. We opted to use 22 topics based on these metrics. Note that the results presented are only one of multiple potentially appropriate models and labeling is relatively subjective; topics are better understood by looking at the top words, particularly those more distinctive to the topic. We explored additional models using different algorithms and hyperparameters (e.g., number of topics, within those suggested as potentially appropriate by our tests), and these provided mostly comparable results. Also note that the topic model was run at the post level (i.e., each post was a document), not weighted by blog. The analysis used variational EM, but Gibbs sampling performed similarly. The LDA was run in Python, using Gensim's package (Rehurek & Sojka, 2010). We used the interactive visualization (available in online repository) to interpret topics because it allows for variation of relevance of words within topics. In other words, it allows users to explore words that occur more or less distinctively within each topic. This aids in interpretation. For a simple table of the top 10 words per topic, see Table S1.

LDA provides the proportion of words in the post that were allocated to each topic, which was used for models with topics as predictors. For analyses with topics as outcomes, we transformed the topic proportions into the count of words for each topic by multiplying the topic proportions per post provided by LDA by the number of preprocessed words per post. This approach still resulted in decimals (due to minor estimation constraints in the LDA model, e.g., the removal of words that appear in only one post), so we rounded down all topic word counts (running the models without rounding resulted in negligible differences). This rounded count of words for each topic was the variable used for models with topic outcomes, as it allowed us to use negative binomial models, which provided easily interpretable rate ratios, and converged much faster than unrounded models.

Results

For a list of the blogs included and some additional information about them, see Table S2.

Throughout the main text and the supplement, models are estimated using the R package lme4 (Bates, Maechler, Bolker, & Walker, 2015); estimates for categorical predictors are least squares means obtained with the R package lsmeans (Lenth, 2016). Analyses using date of the post exclude periods where there was only one active blog (10/2004-9/2008).

As indicated in the main text (footnote 2), we used negative binomial models for count outcomes (e.g., number of comments), which often fitted better than Poisson models due to over-dispersion. For models with topic outcomes, we added the log of the length of the post (more specifically, the number of words in the post that were used for LDA, excluding stopwords, etc.), as an offset. In addition, models with topic outcomes excluded posts for which the LDA post length was zero (e.g., because all words in the post were stopwords), resulting in the removal of 69 posts. For convergence, we used zero-inflated negative binomial models for social media impact outcomes. We used logistic regressions for dichotomous outcomes (e.g., whether a name was mentioned or not).

For judge-nominated names matched to all data: given the impact on date of first mention of matching a nominated name to the wrong person (e.g., due to common last names), post matching was manually checked up by the researchers until the first verified mention and mentions before this date were deleted. In Table S4 we present results using these revised results. However, since the same checking was not done on comment mentions, they are no longer comparable to post mentions. See Table S5 for an uncleaned version of post mentions for comparisons to comment mentions.

Tables Complementing Main Results

See Table S3 for in-depth descriptive statistics. See Tables S4 and S5 for additional information about name nominations and results. See Table S6 for additional information (e.g., p-values) for models from the main text. See Table S7 for date analyses model information. See Table S8 for models with count outcomes (from the main text) accounting for post date. See Table S9 for the predictors' intercepts and standard deviations.

Additional Analyses

Sentiment

Sentiment scores (see Table S10) indicate that blogs tended to have average posts and comments with a greater percentage of positive than negative words, consistent with linguistic positivity biases in general (Matlin & Stang, 1978).

Controlling for career stage, male- and female-led blogs had similar percentages of positive words (Rate Ratio = 1.01, 95% CI = [.95, 1.08]), with male-led blogs having a slightly lower number of negative words (*RR* = 0.89, 95% CI = [.95, 1.08]). Controlling for gender, ECR and non-ECR were mostly similar in sentiment (Positive *RR* = 1.00, 95% CI = [.92, 1.08]; Negative *RR* = 1.03, 95% CI = [.90, 1.17]). Posts that mentioned a nominated researcher (vs. not) barely differed on positive sentiment (*RR* = 0.99, 95% CI = [.97, 1.00]), but negative sentiment was slightly higher in those mentioning nominated names (*RR* = 1.06, 95% CI = [1.02, 1.10]). Posts that mentioned non-nominated names showed similar small differences, (Positive *RR* = 0.99, 95% CI = [.97, 1.01]; Negative *RR* = 1.04, 95% CI = [1.00, 1.07]).

Percentages of words related to *statistics* were negatively related to positive sentiment, both within (*RR* = .96, 95% CI = [.95, .96]) and between (*RR* = .92, 95% CI = [.91, .94]) blogs. Percentages of words related to *statistics* also negatively related to negative sentiment (post-level *RR* = .90, 95% CI = [.87, .92]; blog-level *RR* = .92, 95% CI = [.88, .96]) blogs. Percentage of words related to *research findings* was only slightly related to positive sentiment at the post (*RR* = 1.02, 95% CI = [1.02, 1.03]) and blog (*RR* = .99, 95% CI = [.94, 1.04]) levels. Percentage of words related to *research findings* slightly related to negative sentiment at the post level (*RR* = 1.07, 95% CI = [1.06, 1.08]) and the blog-level (*RR* = 1.03, 95% CI = [0.95, 1.13]). Percentages of words related to *replication* topics were very slightly negatively related to positive sentiment (*RR* = .98, 95% CI = [.98, .99]) at the post level, with a smaller association at the blog level (*RR* = 1.00, 95% CI = [.97, 1.02]). Percentages of words related to *replication* topics were slightly positively related to negative sentiment (*RR* = 1.02, 95% CI = [1.01, 1.03]) at the post level, with a smaller, and negative, association at the blog level (*RR* = 0.98, 95% CI = [.93, 1.03]). For *fraud*, differences on positive (post-level *RR* = 1.00, 95% CI = [.99, 1.00]; blog-level *RR* = 1.01, 95% CI = [0.99, 1.03]) and negative (post-level *RR* = 1.00, 95% CI = [.97, 1.03]; blog-level *RR* = .99, 95% CI = [0.95, 1.03]) content were also small.

Date analyses

Name mentions. Time trends indicated that more recent posts within blogs and blogs with more recent average posts both had more name mentions. Thus, although bloggers did not often mention the nominated researchers, they tended to do so more often over time, and this pattern was stronger than for non-nominated names (which could partially reflect the time of collection of nominated names being more recent).

Topics. Time trends indicated that more recent posts within blogs and blogs with more recent average posts both related to more *replication* and *fraud* words, and slightly to more *statistics* words. Trends for *research findings* were small. Thus, newer posts and blogs included more *replication*, *statistics* and *fraud* talk.

Engagement. Time trends indicate that more recent posts within blogs got more comment engagement, while newer blogs (more recent average post dates) received less engagement.

Similarly to comments, newer posts within blogs got more views, while newer blogs received fewer views. The pattern for social media suggested that more recent posts had more impact, and to a much smaller extent, so did more recent blogs. Note that results for social media impact may speak more to changes in social media use (Facebook, Pinterest, StumbleUpon) over time than about blogs.

Citations. Time trends indicated that post citations increased with post-level recency but decreased for newer blogs (i.e., with blog-level recency).

Citations

Men (M = 10965.63, SD = 16769.11, Median = 4265) had more journal and book citations than women (M = 6750, SD = 14607.92, Median = 397). Analyzing as a rate per publications diminished this difference (M = 67.19 vs. 35.53; Medians = 60.28 vs. 25.33).

References

Bates, D., Maechler, 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.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, 363-370. Retrieved from: http://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*(2), 211. doi: 10.1037/0033-295X.114.2.211

Jamieson, K. H. (1999). Civility in the House of Representatives: The 105th Congress (Report No. 26). Philadelphia, PA: The Annenberg Public Policy Center of the University of Pennsylvania. Retrieved from http://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/REP261.pdf

Jurafsky D., & Martin, J. H. (2016). *Speech and language processing* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*. doi:10.18637/jss.v069.i01

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Cambridge, MA: Schenkman.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence, 29*(3), 436-465.

PsychBrief (2017a, January 25). Improving the psychological methods feed. *PsychBrief*, Available at http://psychbrief.com/improving-the-psychological-methods-feed-2/. Accessed June 10, 2017.

Table S1. 22 LDA topics and their top 10 words and probability within each topic.

#	Topic	Words	P	#	Topic	Words	P
1	Stat: Regression	model	0.03	12	Replication	study	0.03
1	Stat: Regression	regression	0.03	12	Replication	effect	0.03
1	Stat: Regression	variable	0.02	12	Replication	result	0.03
1	Stat: Regression	estimate	0.02	12	Replication	power	0.02
1	Stat: Regression	paper	0.02	12	Replication	size	0.02
1	Stat: Regression	question	0.02	12	Replication	replication	0.01
1	Stat: Regression	effect	0.01	12	Replication	test	0.01
1	Stat: Regression	causal	0.01	12	Replication	significant	0.01
1	Stat: Regression	measure	0.01	12	Replication	sample	0.01
1	Stat: Regression	correlation	0.01	12	Replication	researcher	0.01
2	Stat: General	model	0.07	13	Teaching	student	0.04
2	Stat: General	data	0.05	13	Teaching	class	0.02
2	Stat: General	method	0.02	13	Teaching	teach	0.02
2	Stat: General	bayesian	0.02	13	Teaching	school	0.01
2	Stat: General	analysis	0.02	13	Teaching	time	0.01
2	Stat: General	statistic	0.02	13	Teaching	game	0.01
2	Stat: General	statistical	0.02	13	Teaching	good	0.01
2	Stat: General	work	0.01	13	Teaching	give	0.01
2	Stat: General	problem	0.01	13	Teaching	team	0.01
2	Stat: General	fit	0.01	13	Teaching	player	0.01
3	Demographics	state	0.02	14	Stat: Software	code	0.02
3	Demographics	survey	0.02	14	Stat: Software	function	0.01
3	Demographics	income	0.02	14	Stat: Software	program	0.01
3	Demographics	country	0.01	14	Stat: Software	package	0.01
3	Demographics	rich	0.01	14	Stat: Software	user	0.01
3	Demographics	woman	0.01	14	Stat: Software	version	0.01
3	Demographics	white	0.01	14	Stat: Software	graphic	0.01
3	Demographics	rate	0.01	14	Stat: Software	software	0.01
3	Demographics	number	0.01	14	Stat: Software	page	0.01
3	Demographics	people	0.01	14	Stat: Software	bug	0.01
4	Stat: Frequentist	test	0.07	15	Nutrition	eat	0.04
4	Stat: Frequentist	error	0.05	15	Nutrition	animal	0.03

4	Stat: Frequentist	statistical	0.03	15	Nutrition	die	0.02
4	Stat: Frequentist	inference	0.02	15	Nutrition	food	0.02
4	Stat: Frequentist	sample	0.02	15	Nutrition	wansink	0.02
4	Stat: Frequentist	significance	0.02	15	Nutrition	visualize	0.01
4	Stat: Frequentist	frequentist	0.02	15	Nutrition	specie	0.01
4	Stat: Frequentist	likelihood	0.01	15	Nutrition	eats	0.01
4	Stat: Frequentist	experiment	0.01	15	Nutrition	na	0.01
4	Stat: Frequentist	severity	0.01	15	Nutrition	pig	0.01
5	Law	law	0.03	16	Sci Comm	research	0.02
5	Law	case	0.03	16	Sci Comm	paper	0.02
5	Law	crime	0.03	16	Sci Comm	science	0.02
5	Law	court	0.03	16	Sci Comm	journal	0.02
5	Law	judge	0.02	16	Sci Comm	publish	0.01
5	Law	police	0.02	16	Sci Comm	article	0.01
5	Law	evidence	0.01	16	Sci Comm	review	0.01
5	Law	lawyer	0.01	16	Sci Comm	work	0.01
5	Law	penalty	0.01	16	Sci Comm	university	0.01
5	Law	death	0.01	16	Sci Comm	author	0.01
6	Clinical	patient	0.03	17	RF: Experiment	study	0.03
6	Clinical	trial	0.03	17	RF: Experiment	effect	0.02
6	Clinical	treatment	0.03	17	RF: Experiment	group	0.02
6	Clinical	study	0.02	17	RF: Experiment	difference	0.02
6	Clinical	clinical	0.01	17	RF: Experiment	control	0.01
6	Clinical	cancer	0.01	17	RF: Experiment	data	0.01
6	Clinical	intervention	0.01	17	RF: Experiment	result	0.01
6	Clinical	drug	0.01	17	RF: Experiment	report	0.01
6	Clinical	group	0.01	17	RF: Experiment	sample	0.01
6	Clinical	therapy	0.01	17	RF: Experiment	size	0.01
7	Social Sci	republican	0.01	18	RF: Theoretical	science	0.01
7	Social Sci	voter	0.01	18	RF: Theoretical	claim	0.01
7	Social Sci	year	0.01	18	RF: Theoretical	theory	0.01
7	Social Sci	people	0.01	18	RF: Theoretical	statistical	0.01
7	Social Sci	cost	0.01	18	RF: Theoretical	evidence	0.01
7	Social Sci	government	0.01	18	RF: Theoretical	argument	0.01
7	Social Sci	health	0.01	18	RF: Theoretical	point	0.01
7	Social Sci	money	0.01	18	RF: Theoretical	view	0.01
7	Social Sci	decision	0.01	18	RF: Theoretical	issue	0.01
7	Social Sci	risk	0.01	18	RF: Theoretical	statistic	0.01
8	Fraud	data	0.12	19	Clinical	child	0.02
8	Fraud	analysis	0.02	19	Clinical	people	0.01
8	Fraud	information	0.01	19	Clinical	disorder	0.01
8	Fraud	share	0.01	19	Clinical	mental	0.01
8	Fraud	case	0.01	19	Clinical	disease	0.01

8	Fraud	report	0.01	19	Clinical	parent	0.01
8	Fraud	fraud	0.01	19	Clinical	gene	0.01
8	Fraud	research	0.01	19	Clinical	year	0.01
8	Fraud	researcher	0.01	19	Clinical	autism	0.01
8	Fraud	request	0.01	19	Clinical	study	0.01
9	Stat: Bayesian	probability	0.03	20	Psy: General	word	0.02
9	Stat: Bayesian	prior	0.02	20	Psy: General	language	0.02
9	Stat: Bayesian	distribution	0.02	20	Psy: General	task	0.02
9	Stat: Bayesian	bayesian	0.01	20	Psy: General	seth	0.02
9	Stat: Bayesian	model	0.01	20	Psy: General	cognitive	0.02
9	Stat: Bayesian	hypothesis	0.01	20	Psy: General	personality	0.01
9	Stat: Bayesian	data	0.01	20	Psy: General	memory	0.01
9	Stat: Bayesian	parameter	0.01	20	Psy: General	social	0.01
9	Stat: Bayesian	posterior	0.01	20	Psy: General	psychology	0.01
9	Stat: Bayesian	give	0.01	20	Psy: General	learn	0.01
10	Unidentified	people	0.03	21	Neuroscience	brain	0.02
10	Unidentified	write	0.02	21	Neuroscience	human	0.01
10	Unidentified	thing	0.02	21	Neuroscience	show	0.01
10	Unidentified	good	0.02	21	Neuroscience	time	0.01
10	Unidentified	book	0.02	21	Neuroscience	paper	0.01
10	Unidentified	time	0.01	21	Neuroscience	activity	0.01
10	Unidentified	work	0.01	21	Neuroscience	point	0.01
10	Unidentified	lot	0.01	21	Neuroscience	district	0.01
10	Unidentified	point	0.01	21	Neuroscience	fmri	0.01
10	Unidentified	story	0.01	21	Neuroscience	area	0.01
11	Social Sci	vote	0.05	22	Stat: Figures	graph	0.08
11	Social Sci	election	0.03	22	Stat: Figures	plot	0.05
11	Social Sci	political	0.03	22	Stat: Figures	line	0.03
11	Social Sci	state	0.03	22	Stat: Figures	figure	0.03
11	Social Sci	party	0.02	22	Stat: Figures	table	0.02
11	Social Sci	poll	0.02	22	Stat: Figures	show	0.02
11	Social Sci	democrat	0.02	22	Stat: Figures	data	0.02
11	Social Sci	candidate	0.01	22	Stat: Figures	display	0.02
11	Social Sci	conservative	0.01	22	Stat: Figures	curve	0.01
11	Social Sci	issue	0.01	22	Stat: Figures	log	0.01

Table S2. Blog sample

Blog	Blog Address	Lead Author(s)	Weeks since creation	Number of posts	Posting rate
[citation needed]	talyarkoni.org/blog/	Tal Yarkoni	392.43	166	0.42
Absolutely Maybe	blogs.plos.org/absolutely-maybe/	Hilda Bastian	194.86	103	0.53
Approaching Significance	approachingblog.wordpress.com	Roger Giner-Sorolla	62.29	8	0.13
Basic Statistics	garstats.wordpress.com/	Guillaume Rousselet	56.86	16	0.28
BishopBlog	deevybee.blogspot.co.uk/	Dorothy Bishop	358.29	186	0.52
CogTales	cogtales.wordpress.com	Christina Bergmann & Sho Tsuji	73.14	40	0.55
Crystal Prison Zone	crystalprisonzone.blogspot.co.uk/	Joe Hilgard	265.57	43	0.16
Data Colada	datacolada.org/	Leif Nelson, Joe Simmons, & Uri Simonsohn	186	58	0.31
Doing Bayesian Data Analysis	doingbayesiandataanalysis.blogspot.co.uk/	John K. Kruschke	312.86	203	0.65
Error Statistics Philosophy	errorstatistics.com/	Deborah G. Mayo	292.43	719	2.46
Felix Schönbrodt	www.nicebread.de/	Felix Schönbrodt	260.71	48	0.18
Funderstorms	funderstorms.wordpress.com/	David Funder	232.57	22	0.09
Im a Bayesian and I do what I want	sites.uci.edu/joachim/	Joachim Vandekerckhove	40	3	0.08
Inattentional Coffee	inattentionalcoffee.wordpress.com/	Katherine Wood	14.14	9	0.64
Invariances	jeffrouder.blogspot.co.uk/	Jeff Rouder	109.71	25	0.23
Lorne Campbell	lornecampbell.org	Lorne Campbell	126.29	25	0.2
Matti Heino	mattiheino.com	Matti Heino	62.43	12	0.19
Mind the Brain	blogs.plos.org/mindthebrain/	James C. Coyne	236.71	157	0.66

My Scholarly Goop	myscholarlygoop.wordpress.com/	Hanne M. Watkins	81.14	37	0.46
NeuroAnaTody	neuroanatody.com/	Ana Todorovic	71.86	16	0.22
Neuroconscience	micahallen.org	Micah Allen	388.71	92	0.24
NeuroNeurotic	neuroneurotic.net/	Sam Schwarzkopf	107.71	49	0.45
Neuroskeptic	blogs.discovermagazine.com/neuroskeptic/	Neuroskeptic	441.43	1180	2.67
Not That Kind of Psychologist	asehelene.wordpress.com	Åse Kvist Innes-Ker	347.57	181	0.52
PIGEE	pigee.wordpress.com	Brent Roberts, R. Chris Fraley	335.57	48	0.14
Psych Networks	psych-networks.com/	Eiko Fried	25.86	22	0.85
PsychBrief	psychbrief.com/	Anonymous	178.57	41	0.23
Quick Thoughts	jcoynester.wordpress.com	James C. Coyne	273.29	150	0.55
R Psychologist	rpsychologist.com/	Kristoffer Magnusson	264	39	0.15
Replicability-Index	replicationindex.wordpress.com/	Ulrich Schimmack	123.14	129	1.05
Sak on Science	sakaluk.wordpress.com/	John Sakaluk	105.14	15	0.14
Sometimes I'm Wrong	sometimesimwrong.typepad.com/	Simine Vazire	162.43	50	0.31
Statistical methods, inference, and open science	richarddmorey.org/	Richard D Morey	165.29	32	0.19
Statistical Modeling, Causal Inference, & Social Science	andrewgelman.com/	Andrew Gelman	652	7211	11.06
The 20% Statistician	daniellakens.blogspot.co.uk/	Daniel Lakens	151.71	70	0.46
The Etz Files	alexanderetz.com	Alex Etz	159.86	36	0.23
The Hardest Science	hardsci.wordpress.com	Sanjay Srivastava	421.43	182	0.43
The NeuroEconomist	theneuroeconomist.com/	Gideon nave	75.57	4	0.05
The Skeptical Scientist	timvanderzee.wordpress.com/	Tim van der Zee	41.86	13	0.31
Xenia Schmalzs Blog	xeniaschmalz.blogspot.co.uk/	Xenia Schmalz	90.57	18	0.2
Zeistgeist: Psychological Experimentation, Cognition, Language, and Academia.	rolfzwaan.blogspot.com	Rolf Zwaan	223.29	81	0.36

Table S3. Blogs and bloggers descriptive statistics

	Mean	SD	Min	Q1	Median	Q3	Max
Number of posts	281.4	1128.6	3.0	22.0	43.0	129.0	7211.0
Posting rate (per week)	0.7	1.7	0.1	0.2	0.3	0.5	11.1
Post length (in words)	1116.0	489.8	390.3	808.3	1016.8	1353.4	2423.8
Comment section length (in words)	622.4	841.6	1.5	163.7	323.5	850.5	4817.2
Avg. Comment length (in words)	112.4	105.3	16.0	73.9	91.3	119.0	694.3
Number of comments	4.6	4.2	0.1	1.8	3.3	5.3	16.4
Number of commenters	2.7	2.2	0.1	1.2	2.1	3.7	9.6
Social media impact	36.3	55.8	0	3.9	16.5	37.6	237.0
Post views	2438.9	2468.6	64.6	574.9	1959.8	3472.1	9354.1
Weeks since creation	199.2	139.9	14.1	81.1	165.3	273.3	652.0
Blog citations (total)	22.8	51.4	0.0	1.0	3.0	30.0	308.0
Blogger citations (total)	9651.3	15966.8	3.0	342.0	3199.5	10462.0	72332.0
Blog citations (per post)	0.2	0.3	0.0	0.0	0.1	0.2	1.7
Blogger citations (per publication)	58.2	66.9	0.3	12.5	43.5	83.1	368.5

Note. Mean, standard deviation, and 5-number summaries at the blog level (averaging across posts for post-level variables).

Table S4. Nominated researchers' names, coded gender, number of judges naming the researcher (consensus), number of posts and comments mentioning each researcher, date of first post mention, and post mention rate.

Name	Gender	Consensus	Post mentions	Comment mentions	Date of first mention	Post mentions rate
Bargh	M	8	49	39	3/12/2012	0.185
Baumeister	M	6	28	24	6/4/2010	0.078
Bem	M	2	154	130	1/10/2011	0.472
Bianchi	F	1	4	3	7/16/2012	0.016
Cosmides	F	1	5	6	4/8/2009	0.012
Cote	M	1	1	0	10/26/2016	0.042
Cuddy	F	13	28	71	7/8/2013	0.143
De la o	F	1	0	0	NA	NA
Decelles	F	1	2	4	5/3/2016	0.041
Dijksterhuis	M	3	7	3	5/1/2013	0.034
Duckworth	F	2	5	10	11/16/2010	0.015
Durante	F	3	13	9	5/17/2013	0.064
Dweck	F	2	12	11	12/7/2007	0.025
Finkel	M	1	10	6	2/26/2015	0.090
Fiske	F	5	41	56	4/19/2007	0.079
Förster	M	4	37	36	05/01/2014	0.241
Fredrickson	F	2	20	21	1/23/2013	0.091
Gabriel	F	1	0	18	NA	NA
Gilbert	M	4	55	59	03/24/2010	0.150
Goffman	F	1	2	5	01/20/2016	0.031
Haselton	F	2	2	2	7/31/2013	0.010
Hershfield	M	1	1	2	11/24/2014	0.008
Liberman	F	1	5	10	06/11/2015	0.052
Norton	M	1	18	16	12/13/2007	0.037
Saperstein	F	1	2	2	6/7/2016	0.045
Schnall	F	10	20	17	5/24/2014	0.133
Schwarz	M	6	19	15	10/02/2006	0.035
Seligman	M	1	9	11	7/16/2013	0.046
Smeesters	M	2	25	15	4/6/2012	0.096
Strack	M	8	16	2	11/28/2012	0.070
Stapel	M	1	110	96	9/12/2011	0.378
Stroebe	M	1	7	2	12/6/2012	0.031
Tracy	F	2	13	35	07/31/2013	0.067
Van Bavel	M	4	3	1	7/13/2016	0.077
Vohs	F	5	17	13	3/18/2013	0.080
Wansink	M	3	20	30	2/21/2016	0.337
Willer	M	1	1	5	10/26/2016	0.042
Zhong	M	1	3	2	4/8/2014	0.019

Note. Consensus and mentions were positively correlated (Posts $r = .16$, comments $r = .27$).

Table S5. Nominated researchers mentions

	Mean	SD	Min	Q1	Median	Q3	Max
% All – Posts	0.11	0.09	0.00	0.03	0.10	0.16	0.34
% All – Comments	0.05	0.06	0.00	0.00	0.03	0.08	0.22
# All – Posts	14.8	38.8	0.0	1.0	4.0	11.0	240.0
# All – Comments	14.3	53.9	0.0	0.0	2.0	5.3	323.0
# Individual – Posts	21.3	29.8	0	5.5	14	22.5	154
# Individual – Comments	20.2	27.9	0	3	10	22.5	130

Note. Mean, standard deviation, and five-number summary of nominated researcher mentions. Rows calculated for all researchers indicate the blog-level summary of the number/percentage of posts/comments that mention at least one nominated researcher. Rows calculated for individual researchers is a researcher-level summary (i.e., each researcher is an observation), and summarizes the total number of posts mentioning individual nominated researchers.

Table S6. Additional post-level model information.

Outcome	Predictor	Intercept	Estimate	2.5% CI	97.5% CI	Statistic	P-value	Intercept SD	Estimate SD
Nominated name	Gender	0.09	1.49	0.89	2.48	1.51	0.131	0.72	
Nominated name	Career-stage	0.09	1.14	0.65	2.01	0.45	0.652	0.72	
Non-nominated name	Gender	3.28	2.42	1.32	4.45	2.84	0.005	0.98	
Non-nominated name	Career-stage	3.28	1.49	0.81	2.74	1.29	0.196	0.98	
Statistics	Gender	0.15	1.14	0.85	1.53	0.88	0.380	0.80	
Statistics	Career-stage	0.15	0.71	0.50	1.02	-1.87	0.062	0.80	
Statistics	Nominated name	0.13	0.79	0.73	0.86	-5.72	<.001	0.84	
Statistics	Non-nominated name	0.13	1.12	1.05	1.19	3.48	<.001	0.84	
Research findings	Gender	0.1	1.22	0.94	1.59	1.52	0.130	0.38	
Research findings	Career-stage	0.1	0.84	0.65	1.09	-1.30	0.192	0.38	
Research findings	Nominated name	0.08	1.29	1.19	1.40	6.01	<.001	0.34	
Research findings	Non-nominated name	0.08	1.35	1.26	1.44	8.36	<.001	0.34	
Replication	Gender	0.09	2.00	1.17	3.41	2.54	0.011	0.86	0.08
Replication	Career-stage	0.09	0.97	0.61	1.56	-0.11	0.916	0.86	
Replication	Nominated name	0.09	2.00	1.76	2.28	10.36	<.001	0.78	
Replication	Non-nominated name	0.09	1.38	1.24	1.54	5.69	<.001	0.78	
Fraud	Gender	0.02	1.80	1.19	2.73	2.77	0.006	0.54	
Fraud	Career-stage	0.02	1.09	0.71	1.67	0.39	0.700	0.54	
Fraud	Nominated name	0.03	1.58	1.33	1.88	5.14	<.001	0.57	
Fraud	Non-nominated name	0.03	0.92	0.79	1.06	-1.14	0.254	0.57	
Comments	Gender	0.99	2.28	1.57	3.31	4.34	<.001	0.95	
Comments	Career-stage	0.99	2.16	1.37	3.4	3.33	<.001	0.95	
Comments	Statistics (g)	2.71	1.01	0.88	1.15	0.09	0.931	1.15	0.3
Comments	Statistics (m)	2.71	1.07	0.82	1.39	0.48	0.632	1.15	
Comments	Research Findings (g)	2.69	1.01	0.91	1.13	0.2	0.838	1.14	0.19
Comments	Research Findings (m)	2.69	1.15	0.74	1.77	0.63	0.532	1.14	
Comments	Replication (g)	2.66	1.17	1.05	1.3	2.79	0.005	1.15	0.22
Comments	Replication (m)	2.66	1	0.78	1.27	-0.03	0.979	1.15	

Comments	Fraud (g)	3.12	0.96	0.88	1.06	-0.82	0.411	1.12	0.12
Comments	Fraud (m)	3.12	0.89	0.75	1.05	-1.38	0.167	1.12	
Comments	Nominated name	1.73	1.96	1.78	2.16	13.76	<.001	1.17	
Comments	Non-nominated name	1.73	1.54	1.43	1.65	11.61	<.001	1.17	
Commenter	Gender	0.77	1.99	1.45	2.73	4.25	<.001	0.81	
Commenter	Career-stage	0.77	1.96	1.33	2.89	3.41	0.001	0.81	
Commenter	Statistics (g)	1.82	0.98	0.88	1.09	-0.42	0.678	0.99	0.23
Commenter	Statistics (m)	1.82	1.01	0.8	1.28	0.11	0.91	0.99	
Commenter	Research Findings (g)	1.86	1.1	1.08	1.12	8.37	<.001	0.98	
Commenter	Research Findings (m)	1.86	1.04	0.75	1.45	0.24	0.814	0.98	
Commenter	Replication (g)	1.89	1.19	1.17	1.22	16.25	<.001	0.98	
Commenter	Replication (m)	1.89	0.99	0.8	1.21	-0.14	0.89	0.98	
Commenter	Fraud (g)	2	0.97	0.89	1.05	-0.73	0.468	0.97	0.11
Commenter	Fraud (m)	2	0.93	0.8	1.07	-1.04	0.301	0.97	
Commenter	Nominated name	1.34	1.64	1.52	1.78	12.19	<.001	0.99	
Commenter	Non-nominated name	1.34	1.33	1.26	1.42	9.34	<.001	0.99	
SM impact	Gender	1.99	2.45	1.18	5.07	2.41	0.016	2.04	
SM impact	Career-stage	1.99	5.77	2.02	16.47	3.27	0.001	2.04	
SM impact	Statistics (g)	9.26	0.89	0.77	1.04	-1.49	0.136	1.92	0.24
SM impact	Statistics (m)	9.26	0.65	0.43	0.98	-2.08	0.038	1.92	
SM impact	Research Findings (g)	10.84	0.98	0.78	1.23	-0.19	0.85	2.01	0.5
SM impact	Research Findings (m)	10.84	0.82	0.42	1.62	-0.57	0.57	2.01	
SM impact	Replication (g)	6.73	1.41	1.15	1.74	3.22	0.001	2	0.48
SM impact	Replication (m)	6.73	1.32	0.89	1.95	1.37	0.169	2	
SM impact	Fraud (g)	9.53	0.97	0.85	1.09	-0.56	0.58	2	0.12
SM impact	Fraud (m)	9.53	1.14	0.84	1.55	0.86	0.39	2	
SM impact	Nominated name	3.42	1.92	1.29	2.85	3.21	0.001	2.15	0.47
SM impact	Non-nominated name	3.42	3.08	1.86	5.08	4.39	<.001	2.15	0.9
Views	Gender	612.8	2.72	0.94	7.87	1.84	0.065	1.17	
Views	Career-stage	612.8	1.35	0.5	3.65	0.59	0.553	1.17	
Views	Statistics (g)	1296.5	1.16	0.95	1.4	1.47	0.142	1.25	0.41

Views	Statistics (m)	1296.5	1.15	0.74	1.77	0.62	0.538	1.25	
Views	Research Findings (g)	1273.3	1.03	0.89	1.19	0.38	0.707	1.27	0.28
Views	Research Findings (m)	1273.3	1.17	0.68	2.03	0.56	0.574	1.27	
Views	Replication (g)	824.52	1.23	1.14	1.32	5.31	<.001	1.22	
Views	Replication (m)	824.52	1.41	0.95	2.09	1.69	0.091	1.22	
Views	Fraud (g)	1415.4	0.83	0.77	0.89	-5.32	<.001	1.26	
Views	Fraud (m)	1415.4	0.95	0.74	1.21	-0.43	0.665	1.26	
Views	Nominated name	851.93	1.68	1.35	2.1	4.63	<.001	1.28	
Views	Non-nominated name	851.93	1.54	1.24	1.9	3.92	<.001	1.28	
Post citations	Gender	0.03	2.28	1.07	4.85	2.15	0.032	1.02	
Post citations	Career-stage	0.03	2.52	1.03	6.18	2.02	0.044	1.02	
Post citations	Statistics (g)	0.08	1.11	0.88	1.39	0.87	0.385	1.11	0.12
Post citations	Statistics (m)	0.08	1.08	0.74	1.57	0.39	0.693	1.11	
Post citations	Research Findings (g)	0.08	1.12	1.02	1.24	2.3	0.022	1.12	
Post citations	Research Findings (m)	0.08	0.97	0.63	1.49	-0.15	0.882	1.12	
Post citations	Replication (g)	0.06	1.55	1.43	1.68	11.04	<.001	1.03	
Post citations	Replication (m)	0.06	1.25	0.99	1.58	1.89	0.059	1.03	
Post citations	Fraud (g)	0.09	1	0.9	1.11	0.03	0.979	1.07	
Post citations	Fraud (m)	0.09	0.87	0.7	1.07	-1.34	0.182	1.07	
Post citations	Nominated name	0.04	2.95	2.22	3.93	7.43	<.001	1.14	
Post citations	Non-nominated name	0.04	1.71	1.23	2.38	3.18	0.002	1.14	

Note. All test statistics are z-values. Models using gender and ECR control for each other. Categorical predictors were dummy-coded (Female, ECR, and no mention as baseline), post-level continuous predictors were standardized at the post-level (g) and blog-level (m), controlling for each other. Intercept is in the response scale, estimates are rate ratios. Intercept random effect standard deviation is provided. If model converged, random slope was added, and its standard deviation is reported. Models were analyzed using negative binomial mixed effects regressions, except for name mentions outcomes, which were analyzed using logistic mixed effects regressions. Social media impact accounted for zero-inflation for convergence. Topics were analyzed using an offset for log of post length, and 69 posts were removed due to zero preprocessed post lengths. See Table S9 for predictor intercept and SD.

Table S7. Date analyses

Outcome	Predictor	Intercept	Estimate	2.50%	97.50%	Statistic	P-value	Intercept SD	Estimate SD
SM impact	Date (g)	2.9	3.15	2.33	4.27	7.39	<.001	2.17	0.86
SM impact	Date (m)	2.9	1.44	0.89	2.32	1.48	.140	2.17	
Views	Date (g)	2132.6	1.52	1.26	1.83	4.45	<.001	1.32	0.39
Views	Date (m)	2132.6	0.73	0.54	0.98	-2.11	.035	1.32	0.39
Post citations	Date (g)	0.11	1.21	1	1.47	1.95	.051	1.22	0.32
Post citations	Date (m)	0.11	0.71	0.52	0.96	-2.25	.024	1.22	
Comments	Date (g)	5.53	1.17	1.02	1.34	2.31	.021	1.09	0.34
Comments	Date (m)	5.53	0.64	0.49	0.84	-3.25	.001	1.09	
Commenter	Date (g)	3.39	1.17	1.05	1.31	2.78	.005	0.91	0.27
Commenter	Date (m)	3.39	0.69	0.55	0.85	-3.43	<.001	0.91	
Statistics	Date (g)	0.13	1.02	0.98	1.07	0.97	.330	0.84	0.08
Statistics	Date (m)	0.13	1.05	0.87	1.26	0.49	.621	0.84	
Research findings	Date (g)	0.11	1.05	1.01	1.10	2.39	.017	0.34	0.05
Research findings	Date (m)	0.11	1.00	0.92	1.08	-0.06	.954	0.34	
Replication	Date (g)	0.10	1.41	1.36	1.46	17.64	<.001	0.74	
Replication	Date (m)	0.10	1.24	1.04	1.47	2.41	.016	0.74	
Fraud	Date (g)	0.02	1.19	1.14	1.26	6.97	<.001	0.54	
Fraud	Date (m)	0.02	1.08	0.94	1.24	1.05	.292	0.54	
Nominated name	Date (g)	0.09	1.35	1.15	1.57	3.78	.001	0.68	0.26
Nominated name	Date (m)	0.09	1.22	1.02	1.44	2.23	.026	0.68	
Non-nomin. name	Date (g)	6.19	1.27	1.08	1.51	2.8	.005	0.9	0.3
Non-nomin. name	Date (m)	6.19	1.15	0.91	1.46	1.19	.235	0.9	

Note. All model specifications are the same as in Table S6. Interpret date analyses inferentials with additional caution, as the models do not account for potential internal structure (e.g., autocorrelation).

Table S8. Engagement and citations models controlling for post date

Outcome	Predictor	Intercept	Date	Estimate	Statistic	P-value
SM impact	Gender	0.55	8.69	1.21	0.52	.600
SM impact	ECR	0.55	8.69	2.58	1.55	.120
SM impact	Statistics (g)	0.76	0.1	0.97	-1.07	.285
SM impact	Statistics (m)	0.76	0.1	0.59	-2.35	.019
SM impact	Research Findings (g)	1	0.1	0.91	-0.83	.410
SM impact	Research Findings (m)	1	0.1	0.79	-0.65	.520
SM impact	Replication (g)	0.88	0.11	1.29	11.14	<.001
SM impact	Replication (m)	0.88	0.11	1.13	0.58	.560
SM impact	Fraud (g)	2.62	0.11	0.84	-2.46	.014
SM impact	Fraud (m)	2.62	0.11	0.71	-1.67	.095
SM impact	Nominated name	0.38	0.11	1.63	5.87	<.001
SM impact	Non-nominated name	0.38	0.11	2.57	3.78	<.001
Views	Gender	143.8	0.32	2.83	1.69	.090
Views	ECR	143.8	0.32	2.62	1.7	.090
Views	Statistics (g)	403.8	0.33	1.16	1.52	.128
Views	Statistics (m)	403.8	0.33	0.97	-0.13	.900
Views	Research Findings (g)	461.44	0.35	1.06	0.94	.349
Views	Research Findings (m)	461.44	0.35	1.12	0.33	.744
Views	Replication (g)	291.59	0.3	1.16	2.03	.043
Views	Replication (m)	291.59	0.3	1.26	0.99	.325
Views	Fraud (g)	483.2	0.31	0.8	-6.77	<.001
Views	Fraud (m)	483.2	0.31	0.9	-0.73	.470
Views	Nominated name	315	0.33	1.36	2.88	.004
Views	Non-nominated name	315	0.33	1.46	3.73	<.001
Post citations	Gender	0.02	0.67	2.19	2.03	.043
Post citations	ECR	0.02	0.67	2.86	2.25	.024
Post citations	Statistics (g)	0.05	0.62	1.03	0.26	.800
Post citations	Statistics (m)	0.05	0.62	0.99	-0.05	.960
Post citations	Research Findings (g)	0.06	0.65	1.1	1.87	.061
Post citations	Research Findings (m)	0.06	0.65	0.96	-0.19	.847
Post citations	Replication (g)	0.05	0.75	1.48	9.83	<.001
Post citations	Replication (m)	0.05	0.75	1.22	1.61	.110
Post citations	Fraud (g)	0.06	0.65	0.98	-0.44	.657
Post citations	Fraud (m)	0.06	0.65	0.84	-1.56	.119
Post citations	Nominated name	0.03	0.7	2.64	6.68	<.001
Post citations	Non-nominated name	0.03	0.7	1.59	2.75	.006
Comments	Gender	1.22	0.49	1.57	2.27	.023
Comments	ECR	1.22	0.49	0.84	-0.69	.491
Comments	Statistics (g)	1.45	0.5	1	0.06	.949
Comments	Statistics (m)	1.45	0.5	1.04	0.24	.812
Comments	Research Findings (g)	1.45	0.5	0.99	-0.18	.857
Comments	Research Findings (m)	1.45	0.5	1.17	0.63	.530

Comments	Replication (g)	1.56	0.52	1.12	1.93	.054
Comments	Replication (m)	1.56	0.52	0.95	-0.39	.697
Comments	Fraud (g)	1.75	0.5	0.99	-0.4	.692
Comments	Fraud (m)	1.75	0.5	0.86	-1.38	.169
Comments	Nominated name	1.15	0.51	1.52	9.34	<.001
Comments	Non-nominated name	1.15	0.51	1.29	7.61	<.001
Commenter	Gender	0.85	0.6	1.47	2.36	.018
Commenter	ECR	0.85	0.6	1.12	0.53	.595
Commenter	Statistics (g)	1.16	0.61	0.97	-0.48	.629
Commenter	Statistics (m)	1.16	0.61	0.99	-0.06	.955
Commenter	Research Findings (g)	1.19	0.61	1.06	6.13	<.001
Commenter	Research Findings (m)	1.19	0.61	1.03	0.14	.887
Commenter	Replication (g)	1.3	0.62	1.09	8.63	<.001
Commenter	Replication (m)	1.3	0.62	0.95	-0.47	.638
Commenter	Fraud (g)	1.31	0.6	0.97	-0.86	.390
Commenter	Fraud (m)	1.31	0.6	0.9	-1.21	.225
Commenter	Nominated name	0.99	0.62	1.36	7.99	<.001
Commenter	Non-nominated name	0.99	0.62	1.2	6.2	<.001

Note. For social media impact, year (instead of the full date) was used as a control, due to convergence.

Table S9. Predictors' means and SDs

Variable	Predictor Mean	Predictor SD
Statistics (g)	20.4	22.5
Statistics (m)	20.4	16.3
Research Findings (g)	11.8	10.3
Research Findings (m)	11.8	4.8
Replication (g)	18.9	10.7
Replication (m)	18.9	14.4
Fraud (g)	3.26	5.3
Fraud (m)	3.26	2.2
Date (g)	12/17/2014	811 (days)
Date (m)	12/17/2014	532 (days)

Note. Predictors' means and standard deviations are obtained from an unconditional mixed model (i.e., with the predictors as outcomes in an intercept-only model).

Table S10. Sentiment analysis

Sentiment	Mean	SD	Min	Q1	Median	Q3	Max
Post: Positive	.15	.02	.10	.13	.15	.16	.20
Post: Negative	.07	.02	.03	.06	.07	.08	.11
Comments: Positive	.16	.04	.11	.14	.15	.17	.34
Comments: Negative	.07	.02	.04	.06	.07	.08	.11

Note. Blog-level mean, standard deviation, and 5-number summary for percentage of positive/negative words per post/comment.