

Exploring Research Methods Blogs in Psychology:
Who Posts What about Whom, with What Effect

Gandalf Nicolas^a

Xuechunzi Bai

Susan T. Fiske

Department of Psychology, Princeton University, Princeton NJ 08544

In press, *Perspectives on Psychological Science*

Author Note

^aCorresponding author: Department of Psychology, Princeton University

330 Peretsman-Scully Hall, Princeton, NJ 08540

Email: gnf@princeton.edu

Supplement, data and code available at:

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

Abstract

During the methods crisis in psychology and other sciences, much discussion developed online, in forums such as blogs and other social media. Hence, this increasingly popular communication channel of scientific discussion itself needs exploration, to inform current controversies, record the historical moment, improve methods communication, and address equity issues. Who posts what about whom, with what effect? Does a particular generation or gender contribute most? Do blogs focus narrowly on methods, or do they cover a range of issues? How do they discuss individual researchers, and how do readers respond? What are some impacts? Web scraping and text analysis techniques provide a snapshot characterizing 41 current research methods blogs in psychology. Bloggers mostly represented psychology's traditional leaderships' demographic categories: primarily male, mid-to-late career, associated with American institutions, and with established citation counts. As methods blogs, their posts do mainly concern statistics, replication (particularly statistical power), and research findings. The few posts that mentioned individual researchers substantially focused on replication issues; they received more views, social media impact, comments, and citations. Male individual researchers were mentioned much more often than women. Further data can inform perspectives about these new channels of scientific communication, with the shared aim of improving scientific practices.

Keywords: research methods, blogs, social media, natural language processing, replicability

Exploring Research Methods Blogs in Psychology: Who Posts What about Whom, with What Effect

Science is renewed by periodic disruptions, and the current methods crisis appears to be no exception to this abrupt form of self-correction. Besides gradual types of feedback—peer review, replication and extension, competitive model-testing—more urgent paradigm shifts also have a role in self-correction. For example, in the late 1970s social psychology underwent a scientific crisis arguably worsened by incomplete methods reporting and inadequate operationalization (Greenwald, 1976). Methodological reforms suggested reporting effect sizes and conducting power analyses (Rosenthal, 1979). Critics also identified the crisis as having theoretical roots (Deutsch, 1976; Smith, 1978).

Psychology's current crisis revolves more pointedly around improving methods. The field is debating false-positive results, neglect of null findings, confirmation bias in data analysis, and incomplete reporting, among others (Nelson, Simmons, & Simonsohn, 2018). Questionable research practices, unsophisticated power analyses, and neglect or misuse of meta-analyses add to the issues discussed (Shrout & Rodgers, 2018). Proposed remedies often include open science norms (Munafo et al., 2017): preregistering hypotheses and analyses, more complete reporting, posting data and code, and more open peer review.

The current crisis in psychology also stands out in a different way. Much of the discussion surrounding these topics seems to have developed online, as social media have become popular forums for research methods and meta-science. For example, on Facebook, two general psychology methods discussion groups were created in 2015 (perhaps the first of their kind), the largest of which has over 12,000 members. Many of the authors at the forefront of the methods discussion maintain blogs (e.g., Nelson, Simmons, & Simonsohn, 2018). One blogger's post reportedly had more views than the journal article to which it was a response (Lakens,

2017). However, beyond these informal observations, few sources report usage trends in scientific social media. Thus, here we provide an initial exploration of the author demographics, properties, content, and impact of one of these online forums: research methods blogs.

Blogs have the potential to reach large audiences, promote discussion, and shape scientific norms faster than traditional channels. However, some have voiced concerns about the current online discourse being often unmoderated, citing for example that online criticism of an individual researcher across social media can become excessive (Fiske, 2017). Others have spoken out against the “(relatively infrequent) occasions when disagreements erupt into personal attacks and harassment” (Coan et al., 2017), both online and offline. Some have argued that these behaviors could lead to intolerance for new research ideas (Sabeti, 2018), while others have speculated that they relate to researchers’ feelings of disengagement with the field (West & Skitka, 2017). Other much-discussed issues include blogger diversity (e.g., in terms of race, gender): Some point out blogs’ ability to provide everyone with a voice (e.g., Lakens, 2017), while others discuss demographic discrepancies in who participates (e.g., West & Skitka, 2017; see next sections). Of course, most of these concerns apply to more traditional scientific communication venues, such as peer reviews and conferences. However, social media provide an accessible format to study these issues in a quantifiable manner. Nonetheless, we caution throughout against drawing contrasting inferences between scientific forums from our results alone. This paper is exploratory.

Why Blogs?

Among social media, we focus on blogs for several reasons, all supported by data (see Results). First, blogs have appeared as scientific forums for much longer than other outlets such as Facebook, providing data over several years, and making them a potentially more established medium. Second, posts are longer than other social media, allowing for a more in-depth analysis.

Finally, several other minor concerns favored using blogs over other data-source options; for example, blogs allow the possibility of collecting variables such as post citations.

Documenting the features of psychology methods blogs matters to our science now, to constructively address several issues: (a) Describing how the blogs approach improving psychological *methods* can inform both the blogs' continuing contributions and the focus of methods development. (b) Recording a slice of current *history*, such as documenting some of blogs' basic features now, will facilitate later comparisons to current trends. (c) Previously discussed ongoing *controversies* about the blogs may benefit from evidence, such as some provided here. (d) Finally, evidence can inform some *equity* concerns about who blogs, what is blogged, about whom, and with what effect. In each case, the current exploration is potentially useful because a priori the results are not obvious; indeed, the authors disagreed about what we might find, and this is new territory, so we opted for an exploratory approach.

Who Blogs?

To elaborate regarding who blogs, one perspective could be that, because blogs afford democratic open-access to voice, we might expect better demographic representation; a contrasting view is that traditional power structures would reproduce in this new medium. For gender, male bloggers could predominate, for example because of traditional gender patterns—in emergent leaders (e.g., Eagly & Karau, 1991) or in assertiveness (Feingold, 1994). Another possibility is that women blog more than men because it might fit gendered leadership roles (e.g., democratic, participatory, communal leadership; Eagly & Karau, 1991) or due to increased numbers of female academics (e.g., in the US: Burrelli, 2008; National Science Foundation, 2015). Similarly, blogger career-stage would be hard to predict. As a newer digital medium, younger bloggers might predominate. Alternatively, more established researchers might enjoy a larger reach and thus engage in blogging more often. Given blogs' role in scientific

communication, a lack of diverse voices is detrimental for equity and maybe minorities' attitudes toward the medium, as well as its future adoption patterns (e.g., Stout et al., 2011).

Blogs What?

Regarding what the blogs say, the controversies described earlier represent two contrasting expectations, at the polarized extremes. One perspective is that the blogs contain entirely technical topics (e.g., statistics). The contrasting view is that the blogs mainly discuss specific individuals' research. Both perspectives are unrealistic prototypes, but they illustrate the endpoints of a potential distribution of topics. Additionally, content analysis describes the historical moment and how blogs are approaching methods improvement (e.g., which topics, such as meta-analyses or statistical power, do blogs emphasize, and which receive attention?).

Blogs about Whom?

Coverage of researchers as the focus of criticism could inform controversies and equity concerns. For example, women score lower on some metrics of influence (Diener, Oishi, & Park, 2014). Thus, men could be mentioned more often to the extent that more publications and impact are a factor in social media post-publication review. Alternatively, this same power imbalance, as well as gender bias, could lead to higher rates of mentioning female researchers (e.g., Elsesser, 2018). Women might also be salient because they are more cited per article (in at least one flagship journal; Cikara, Rudman, & Fiske, 2012).

Blogs with What Effects?

Finally, several possibilities emerge for the impact of blogging. For example, gender and career-stage differences could reflect the just-noted patterns in more traditional outlets (e.g., journal publications). Alternatively, blogs may provide a democratized space that allows different demographics to have more equal reach (Lakens, 2017).

Overall, as noted, analyzing blogs could inform how to improve methods, describe current historical trends, bring data to controversies, and address equity concerns. These are among the motivations for exploring the who, what, whom, and effects of methods blogs.

Precedents for Who Posts What about Whom, with What Effects?

A few scattered studies are beginning to describe social media's roles in psychology. For example, in a sample of speakers at the 2016 Psychonomic Society conference (N = 327), 19% had Twitter accounts and 6% had a blog (Weinstein & Sumeracki, 2017). These findings suggest a passive use of social media: Few are leading the conversation, but many more may be following it. Additionally, women in the sample tended to have Twitter accounts more often than men. As another example, a short demographic analysis of bloggers (described below) appeared, but it did not further explore characteristics of their blogs (PsychBrief, 2017a).

Others have described participants' self-reports about platforms besides blogs. For example, a survey of members of the Society for Personality and Social Psychology listserv found that men reported using the groups more often; the most common reported uses were reading about and sharing new research findings and methods, as well as learning academic gossip (West & Skitka, 2017). Additionally, a non-trivial number of respondents reported feeling personally attacked in the groups (or Twitter) at least once (54% said never, 12% sometimes).

Overview

Given the lack of data on the characteristics of psychology's social media (especially blogs), the discussions that have surrounded these forums, and their apparently increasing role in setting scientific norms, the current paper aimed to provide additional insights about research methods blogs in psychology. As new forms of communication grow in popularity, understanding patterns of adoption, content, and impact might provide insights into how to maximize their positive influence toward improving our science. We explore who is posting,

what they are posting, who they are mentioning, and how people are engaging with them. We use web scraping and natural language processing to describe some properties of a sample of current blogs, including their general characteristics (e.g., frequency of posts), bloggers (e.g., demographics), content (e.g., topics covered), and impact (e.g., post citations).

Reported results are strictly exploratory (*p*-values are not reported in the main text but are included in the Supplement's Table S6 for interested parties; uncertainty measures are provided only as an exhortation that estimates not be interpreted as overly precise, not for inferential purposes). All data and code are available in the online repository.

Method

Materials and Procedures

This section describes how we obtained the blog sample as well as how we operationalize who is blogging (blogger characteristics), what they are posting (topics), about whom (researcher mentions), and with what effects (engagement and citations).

Blog sample. A list of prominent, active research methods blogs in psychology and related disciplines appears in PsychBrief's blog feed (2017b), which included 43 blogs at the time of data collection (by April 11, 2017). However, only 41 blogs (11,539 posts) were useable because one of the links listed was a journal, and another's server was unstable, so we were unable to access it reliably. The blog feed did cover a sizable number of research methods blogs in the area at the time of data collection, and the source was independent from the authors' judgment. We also sought suggestions from some experts in the field, all of which were included in the feed. Nonetheless, we are aware of at least 5 blogs at the time of collection that could meet the criteria of being "research methods blogs in psychology" that were not included in the feed at collection time, and other readers would surely be aware of more. As with any convenience sampling, we note that this sample is incomplete and biased, including toward salience or other

preferences the feed author might have (although note the author of the feed had actively attempted to minimize gender, race, and career-stage bias by the time of data collection; PsychBrief, 2017a). See Supplemental Materials for a description of web scraping and preprocessing.

Some computed variables describe blog characteristics, including the number of posts, date of blog creation, and posting rate (number of posts divided by weeks since date of creation). For analyses, these are blog-level variables (i.e., they do not vary within blogs). Variables in the following sections may vary between posts of the same blog (i.e., they are at the post-level).

Blogger characteristics: Who? There were 70 total identifiable bloggers across the 41 blogs, due to 5 multi-author blogs. Bloggers' demographic data (gender, nation of institutional affiliation, and career-stage) were obtained from various sources. We used institutional websites such as the blogger's or university website, or reputable journalistic sources to code gender based on pronouns. Thus, possibly, our coding may not match some of the bloggers' identities. Missing values remained, particularly for secondary authors in multi-authored blogs, but also for four primary authors (two of whom were anonymous bloggers).

Early career researcher (ECR) was defined as someone who has been hired to perform research within the past 5 years (Research Excellence Framework, 2014); we included graduate students and postdocs in this category, and due to the time-changing nature of this coding, categorized as ECRs bloggers who met the criteria for most of their posts. ECR determinations were based on the bloggers' websites and CVs. In addition, to provide some insight into the geographic distribution of blogs, we coded the country of institutional affiliation of the bloggers from sources such as Google Scholar.

For analyses, gender was further recoded into male or female and career-stage into ECR or not-ECR. Guest authors are uncommon for most blogs and are not included in the analyses;

specific authors were only identified when the information was available as metadata. For multivariate analyses we focus on gender and career stage as the most relevant data to current controversies for which we considered to have reliable coding for most bloggers.

For models at the blog-level, we recoded demographics, such that multiple-author blogs with gender-congruent data (e.g., two female bloggers) were coded accordingly (3 blogs); otherwise, if the blog had a lead author (based on the author page, blog metadata, and number of posts), their information was used (2 blogs).

Topics: What? After preprocessing the text, we used Latent Dirichlet Allocation (LDA; Blei, Ng, & Jordan, 2003) to model the topics covered by the posts. LDA is a probabilistic generative model that attempts to discover latent semantics embedded in document collections. LDA assumes that topics can be represented as word distributions, such that each topic is associated to each word from the posts to different extents. For example, one of the topics suggested by the model (which we later labeled to be about *teaching*) had high probabilities for words such as “student” and “class” but low probability for words unrelated to teaching, such as “bayesian” or “republican.”

LDA also assumes that documents (i.e., posts) can be represented as distributions of topics, such that each post can cover a variety of topics to different extents. For example, LDA suggested that a post titled “Bayesian Perspectives on Publication Bias” had high proportions of words assigned to two topics, which we later labeled *Replication* (52%) and *Statistics* (32%), while a post titled “Why brain scanners make your head spin” had over 75% of its words assigned to a topic we labeled *Neuroscience*. Thus, through LDA, we obtained the number and proportion of words within each post that came from the different topics discovered (for more information see Supplement and a comprehensive review in Griffiths, Steyvers, and Tenenbaum, 2007).

Tests to decide on a number of topics (see Supplement) suggested 22 topics. Next, two of the authors independently labeled the topics provided by the 22-topic LDA model. In order to arrive at a label, the authors looked at several indicators, including (a) the top words within each topic, (b) the distinctive words within each topic (e.g., “Stapel”¹ occurred exclusively in the topic labeled as *Fraud*), and (c) example posts with high topic proportions. Thus, labeling took several indicators into account. Nonetheless, topic labels may not fully represent some topics, as multiple labels may sensibly describe topics: labeling prioritized issues of a priori interest (such as *replication* and *fraud*).

Topics interpreted as highly overlapping were combined, resulting in a grouping of 12 final topics, listed in Table 1 (and further detailed in the Supplement, Table S1). For example, for each post we added the number of words allocated to topics such as Bayesian statistics, Frequentist statistics, Regression, Visualization, and other related topics, resulting in an overarching “Statistics” topic word count.

We conduct further analyses on four topics of interest: *statistics* (which included top words such as “model,” “data,” “probability”), *research findings* (e.g., “difference,” “control,” “claim”), *replication* (e.g., “replication,” “effect,” and “power,” related to replicability), and *fraud* (e.g., “fraud,” “share,” “report”; the topic also distinctively included words about retraction, misconduct, and ethics). *Statistics* and *replication* were the most common topics, but we include *research findings* and *fraud* for their potential relevance to other variables (e.g., individual researcher mentions) and to methods improvement.

Researcher mentions: Whom? To provide an initial exploration of bloggers’ discussions of individual researchers, we identified and tagged people’s names in the data, by using an automated method (named entity recognition, essentially a mechanism to locate, classify, and count proper nouns; see Supplement). However, because these names could refer to anyone,

from researchers to statistics named after people (e.g., Cohen's d), we also obtained a list of salient names from judges who have weighed in publicly on the replication crisis, but not as bloggers. All had spontaneously contacted the senior author, implicitly indicating their interest and availability to discuss this topic (in response to Fiske, 2017). Admittedly, judge selection was ad hoc, not blind, and non-random. Automating the identification of researcher last names was not possible (isolated from statistics named after people, statisticians, celebrities, first names, and fragments), and we did not want to trust only our own judgment.

These judges in the field were asked to nominate up to 12 psychological researchers who have been under critical focus, allowing us to obtain a more specific sample of names to compare with general discussions that involve names. Of 16 judges contacted, 13 provided useable responses; 9 were male. The judges generated a list of 38 researchers (20 male). Nominated researchers have been the focus for multiple reasons, and thus this list simply represents salient names that have received post-publication critiques; we make no assessment about the appropriateness of any individual criticism. But these analyses could inform further discussions about post-publication review and provide some correlates of individual mentions.

For analyses on relationships between variables, we matched the names nominated by judges to the named-entity-recognition output. Although this is less reliable than matching to all data, it made an appropriate comparison to non-nominated names (also identified through named entity recognition): When using name mentions as predictors, we simultaneously enter variables on whether a post mentioned a judge-nominated researcher or not, and whether a post mentioned a non-nominated researcher or not. However, for accuracy, we provide descriptive statistics by matching the nominated researchers' last names to the entire corpus of data. Note that some name mentions could possibly refer to another person who shares a name with a nominated researcher (some of these were deleted if identified).

Engagement and citations: What effects? As measures of active engagement, we created variables indicating the number of comments and commenters (i.e., number of comments posted by distinct authors) for each post. Additionally, we collected information on posts' social media impact (including Facebook reactions such as likes, Facebook comments, Facebook shares, as well as the equivalent of likes on Pinterest and StumbleUpon; these variables were aggregated; data from Twitter were not available). As a measure of passive engagement, we requested post views from bloggers, which indicate how many times each post had been accessed. However, due to non-responses, denials to share, or technical issues, we obtained post views for only 59% of the blogs. Thus, an alternative interpretation for all comparisons to post views is that differences are due to missingness. Furthermore, views may be noisy (e.g., because they may indicate either total or unique views). Finally, both post views and social media impact variables were collected in May-June 2018, after most other time-sensitive variables, so care is needed when comparing to other metrics.

As a measure of more long-term impact, we collected data on post citations. We obtained citations by searching the minimal URL for accessing the blog (e.g., by stripping "https://") in Google Scholar. Citations were recorded September 16-21, 2017. We also collected the number of blogger academic publications and citations in Google Scholar for comparison.

Other variables used in analyses are the date and length of the post and the length of the comments to the post. More information about additional variables (e.g., date and sentiment analyses) is offered in the Supplement.

Data Analysis Methods

Data analysis was conducted in R. We first present univariate descriptions of data, then some relationships between variables. Model details appear in the Supplement.² The online

repository provides figures for most results, to display the raw data beyond our models; we also provide the data and code for further exploration.

Over 60% of our data came from a single blog (see Supplement for blog sample information, Table S2), an influence that we attempted to balance by conducting analyses at the post level, accounting for the multilevel structure (i.e., blogs as a random factor) or at the blog level (i.e., with each blog being an observation). Thus, each blog is relatively equally weighted, and most results describe properties of the average blog, not post. For simplicity, we exclude interactions. A maximal random structure (Barr, Levy, Scheepers, & Tily, 2013) was used when models converged.

In general, models allow for outlier influence, but in some cases, we also explored (mostly graphically, in online repository) how the data look without these outliers. Most results are directionally robust to outlier exclusions, except when indicated.

As mentioned, all analyses at the post-level used mixed models to account for non-independence (i.e., posts are nested within blogs). Continuous predictors were standardized (rather than just centered, for convergence and interpretability). Continuous variables with post-level variation used as predictors (mainly posts' topic proportions) were transformed into two variables. First, a post-level variable standardized within each blog (e.g., each post's proportion of *statistics* words, standardized using only the data for its corresponding blog).³ Second, a standardized blog-level variable (e.g., each blog's mean proportion of statistics words across its posts, standardized using all data). This partition allowed us to explore both between-posts and between-blogs variation, controlling for each other (see, e.g., Enders & Tofighi, 2007).

For analyzing date, with blog properties as outcomes, we collapsed across blogs to analyze not only the number of posts by date, but also the number of blogs, and the posting rate as a function of number of blogs. In these, the date predictor was coded with months as units,

higher numbers indicating more recent months. Further date analyses are in the Supplement, Table S7.

Gender and career-stage were always entered simultaneously in models with post-level outcomes. Models estimated (to control for) a date coefficient for engagement and citations outcomes are reported in the Supplement (Table S8) along with other supplementary results. When these models' interpretations differ from the ones presented here (which do not control for date), we report it.

Results

Our exploration addresses several questions: Who? (Does a particular generation or gender contribute most?) What? (Do blogs focus narrowly on methods, or do they cover a range of issues?) Whom? (How do they discuss individual researchers?) Effects? (How do readers respond? What are some impacts?)

Bloggers: Who Posts?

Most bloggers were male (73% of $n = 52$). For comparison base rate, women are now equal or a majority in American psychology (Burrelli, 2008; National Science Foundation, 2015). However, not all bloggers are academics, psychologists, or American—and these figures include clinical training—so maybe not the most appropriate base rates for research methods.

Despite blogs being part of the new media, which might suggest younger authors, the bloggers are not early-career researchers (of $n = 54$, 67% mid-to-late career). Moreover, only 26% of male bloggers were early-career researchers, while 50% of female bloggers were ECRs. This might forecast a generational shift in blogger gender ratios. Bloggers were mostly associated with American universities ($n = 56$, 59%, 13% UK universities, 11% Dutch universities). A previous report on Psychbrief's feed coded most of the lead bloggers as White

(93%; Psychbrief, 2017a). Two of the bloggers were entirely anonymous. In short, most bloggers were established, male, and associated with American institutions.

Blog Properties: Who Posts How Often, and for How Long?

For current context and the historical record, documenting blogs' basic properties is useful (see Table S3 for full details on descriptive statistics). Most blogs had a relatively small number of posts in the 12.5 years since the first blog began (Median = 43), although one prolific blog had 7,211 posts at the time of collection (62.5% of the 11,539 posts analyzed here). Blogs tended to be approximately 3-4 years old. Blogs by female authors were newer (Median weeks old = 126.5, $M = 168.64$, $SD = 124.78$) than male-led blogs (Median weeks old = 165.29, $M = 205.74$, $SD = 144.66$).

Posting rate (i.e., number of posts divided by weeks since creation) was relatively low for most blogs, with a median of a little over 1 post per month (but with outliers averaging up to 11.1 posts per week). Male-led blogs ($M = 0.74$, $SD = 2.08$) had a higher mean post rate than female-led blogs ($M = 0.64$, $SD = 0.66$). However, for medians, because of one male-led outlier, female-led blogs posted more per week (Median = .52) than male-led blogs (Median = .28). More established (non-ECR) bloggers posted at a higher rate ($M = 0.95$, $SD = 2.26$, Median = 0.43) than ECRs ($M = 0.3$, $SD = 0.21$, Median = 0.22).

Time trends indicate that both the overall posting rate (Rate Ratio⁴ = 1.079, 95% CI [1.05, 1.11]) and the blog creation rate ($RR = 1.33$, 95% CI [.97, 1.84]) have increased over the months included. However, the rate of posts per blog decreased over time ($RR = 0.453$, 95% CI [.43, .48]).

The percentage of new blogs led by women per month also grew over the months ($RR = 1.46$, 95% CI [.77, 3.02]). Thus, overall the trends indicate more robust participation by women.

In sum, most blogs were 3-4 years old, posting about monthly, with increases in both new blogs and posts per month. However, the posting rate per blog decreased over time.

Individual Researcher Mentions: Blogging About Whom?

One controversy concerns the rate of posts that mention individual researchers; the data indicate that neither extreme is right on average, but outliers exist. Regarding 38 individual researchers nominated by judges as frequent targets of critical focus, blogs have around 11% of posts mention at least one such researcher. Six blogs mentioned none of the nominated researchers, while one blog went up to 34% of its posts. The percentage of posts mentioning judge-nominated researchers was higher than the proportion of comment sections that did. The most common nominated researcher mentioned differed between blogs, so we collapsed across blogs to look at individual mention patterns. This approach indicated that the most common nominated researcher (even when accounting for date of first mention) was Daryl Bem, who published a paper on precognition (Bem, 2011) that garnered widespread attention (e.g., Engber, 2017). Daryl Bem was mentioned in 154 posts and 130 comment sections. Generally, looked at individually (rather than summing across all names), nominated researchers appeared in a median of 11 posts and 10.5 comment sections. See Supplement (Table S4) for more details.

Part of the controversy also touches on power dynamics: With bloggers being mostly male, are female researchers mentioned most? Female ($M = 3.00$, $SD = 3.38$) and male ($M = 2.95$, $SD = 2.39$) researchers had been mentioned by a similar number of judges on average. However, nominated male researchers were mentioned in over twice as many posts ($M = 28.65$, $SD = 39.23$, Median = 17) as female researchers ($M = 10.61$, $SD = 11.16$, Median = 5) on average. For comments, the difference between male ($M = 24.7$, $SD = 34.49$, Median = 13) and female ($M = 16.27$, $SD = 19.33$, Median = 10) mentions was notable but smaller.

The pattern appears stable over time. Male first mentions are on average 28.9 days more recent than female first mentions. As a rate, differences between male and female nominated researchers remain, with more mentions per week for men ($M = 0.12$, $SD = 0.13$, Median = .07) than women ($M = 0.06$, $SD = 0.04$, Median = .05).

Which bloggers? Men's (vs. women's) posts mentioned nominated researchers more often, Odds Ratio = 1.49, 95% CI = [.89,2.48]. Men also more often mentioned non-nominated named entities (e.g., statisticians, celebrities; $OR = 2.42$, 95% CI = [1.32, 4.45]). Non-ECR versus ECR blogs showed slighter differences (nominated: $OR = 1.14$, 95% CI = [.65, 2.01], non-nominated: $OR = 1.49$, 95% CI = [.81, 2.74]).

Overall, judge-nominated names appeared rarely. There were equal numbers of men and women nominated, but nominated men were mentioned in posts more often. Male bloggers mentioned names more often than female bloggers.

Topics: Posts What?

Regarding blogs' focus on improving methods, topic analysis suggests that blogs' average posts do include a large proportion of words related to *statistics*, *replication*, *research findings*, and *science communication* more broadly (see Table 1). In terms of disciplinary content, none was a particular focus: Blogs covered clinical, neuroscience, and social science topics to similar degrees. Some of these broad topics can subdivide with differing levels of specificity (down to the 22 topics provided by the LDA model), revealing for example that discussions of statistics often revolved around Bayesian statistics. Additionally, the distribution of terms for each topic (see Table S1 in Supplement) suggests that statistical power is the most discussed issue within *replication*-related posts, while in the *fraud* topic, salient words refer to data sharing, and included one nominated researcher's name (Stapel). For results see Table 2.

Which bloggers? Men (vs. women) talked more about *replication*, *fraud*, and *research findings*⁵. ECRs (vs. non-ECRs) talked more about *statistics* and *research findings*.

Mentioning individual researchers? Posts mentioning names, both nominated and not nominated (e.g., statisticians' names), were compared to those mentioning no names. The posts mentioning names had more talk about *replication* and *research findings*. Posts mentioning nominated researchers had a smaller proportion of *statistics* words, while those mentioning non-nominated names had a slightly larger proportion. Posts mentioning nominated researchers had more *fraud*-related words, while posts mentioning non-nominated names had fewer *fraud*-related words.

In sum, the *replication* and *research findings* topics were more common in posts by men or which mentioned nominated researchers. Discussions about *fraud* were more common in posts by men or which mentioned nominated researchers. *Statistics* was covered more by ECRs, and less in posts mentioning nominated researchers.

Engagement: With What Immediate Effects?

In terms of immediate impact through comment engagement, blogs received a median of 3.34 comments per post (2.10 unique commenters), but a prominent blog could receive up to 16.40 comments per post (9.63 from different commenters). Note that these counts include comments by the blog authors themselves; five of the blogs did not allow comments.

For social media impact, blogs received a median of 16.48 interactions (including reactions such as likes, Facebook comments, and shares). Numbers for views were larger, with a median of 1959.81 (min = 64.6, max = 9354.15). Many people may be visiting, but few are engaging.

Which bloggers? More established and male bloggers captured more engagement (See Tables 3 & 4 for results). This was the case for all indicators, including number of comments,

commenters, views, and social media impact, with men mostly obtaining twice or more the size of these indicators. Non-ECRs also tended to double or more ECRs' impact (but when controlling for date of the post ECRs had slightly more comments than non-ECRs).

Mentioning individual researchers? Whether a post mentioned names did relate to engagement: The number of comments in posts that mentioned judge-nominated researchers was almost twice as in posts with no mentions, and it was also somewhat larger in posts mentioning non-nominated names (vs. no mention). The case was similar, but to a smaller degree, for number of commenters. Social media impact and views were also higher for posts mentioning names (but unlike other indicators, non-nominated had a larger effect than nominated names).

Which topics? Increased *replication*-talk at the post-level resulted in increases across all indicators. Blogs that discussed replication more often received more views and social media impact than those that discussed the topic less, but the difference was very small for number of comments and commenters.

The relationship with comments and commenters was small for differences in proportion of *statistics* coverage. On the other hand, social media impact was lower for blogs (and less-so posts) that discussed statistics more often, while post views were slightly higher (except at the blog-level when controlling for date). Thus, statistics posts may be passively engaging, but not as actively discussed or shared.

Fraud-related words slightly related to less engagement across most metrics. Engagement with research-finding-talk depended on the medium, and effects were small.

Overall, views were high, but engagement via comment was relatively low. Engagement was higher when specific names or *replication* appeared. Male and established bloggers elicited more engagement.

Citations: With What Longer-term Effect?

A more formal measure of impact is blog citations in Google Scholar: The same pattern of low numbers of total citations emerged for most blogs (Median = 3, $M = 22.8$, $SD = 51.4$), with exceptions going up to 308 citations. In terms of the bloggers, Google Scholar profiles⁶ include journal articles, chapters, and books but not blog posts, with very few exceptions. All bloggers had some citations (Min = 3, Median = 3199.5, $M = 9651.3$, $SD = 15966.8$), and this varied up to bloggers with citation counts indicating very influential publications (Max = 72,332). The median number of blog citations was 0.07 per post ($M = 0.16$, $SD = .28$, Max = 1.71); the median rate of Google Scholar citations was 43.5 per publication ($M = 58.24$, $SD = 66.87$).

Which bloggers? Men had more post citations than women.⁷ Non-ECRs had more post citations than ECRs. These results are in line with shorter-term impact.

Mentioning specific researchers? Post citations were positively related to name mentions, particularly nominated names which almost tripled the number of citations.

Which topics? In terms of content, *replication* coverage was again the strongest predictor, increasing the post-level citations by over 50%. Post citations were slightly positively related to *statistics* coverage. *Research findings* post-level relationship with citations was positive. At the blog-level, *fraud-talk* was negatively correlated with post citations.

In sum, long-term impact reflected the same variables as did short-term impact: established male bloggers and those discussing specific researchers and *replication* elicited more engagement.

Discussion

Online channels of scientific communication are changing our field's feedback systems. Research methods blogs' data suggest an increasing number of blogs over the period included.

Studying blogs' properties and influence provides avenues to understand how scientific communication is shifting, with opportunities to encourage positive impact and reduce drawbacks. We used text analysis to describe some patterns.

Blogs with What Effect?

Analyses point to a small number of citations for most blogs, but some outliers have more than 25 times as many citations per post as the median blog. For a coarse comparison, the median blogger's journal and book citations per publication was 669 times the number of post citations per post. Similarly, most blogs engaged only around 2 commenters per post (but with potential to get 144 commenters in a post). Patterns were similar (but larger) for engagement through other social media, such as Facebook. Post views, on the other hand, were much larger, with a median of almost 2,000 for blogs with data (but noise, missing data, and differing collection times require caution when comparing across metrics). Thus, influence may be more passive: Readers engage by viewing or sharing the posts, but not necessarily joining a discussion or citing the posts.

We note that some indicators of effects of blogging, such as post views (including both distinct and recurrent views), identities of guest authors, and in some cases even the exact time and date of posting were often unavailable. We exhort bloggers to make as much of this and other information available as possible for future examinations.

Blogs What?

Our results also highlight the diversity of topics covered by methods blogs. Discussions of *replication* were dominated by words such as “power” and “sample size”; other top words, such as “bias” and “report,” might indicate that publication bias and incomplete reporting were also in the conversation. Other topics suggested to remedy the replication crisis, such as meta-analyses or preregistration, seem to be less discussed. The top words of the *fraud* topic suggest it

is closely linked to specific cases (in particular, Stapel) and data sharing. *Replication*-talk related to more impact and was covered more often by men. Relationships for other topics were smaller and less robust, but discussions of the *fraud* topic (mostly covered by men) were associated with less impact. *Statistics* (more often covered by ECRs) related to less social media impact but more passive impact. Thus, less technical discussions of replication-relevant issues seem to draw more attention than topics related to purely statistical discussions.

As an imperfect comparison, we applied the same topic model to 1039 response articles in the journal *Behavioral and Brain Sciences*, from 2014 through 2017 (for a summary of methods related to analyses of this journal, see Supplement). These are open peer commentaries to target articles and may thus be closer to blog posts and peer review than other traditional outlets. From the topics covered by the blogs, the journal articles covered mostly *replication* (13.9%), *science communication* (13.7%), and *theory/research findings* (10.4%). These percentages do not represent actual coverage of these topics and should only be interpreted in relative terms. However, they do suggest differences, such as blogs' higher focus on *statistics*, as well as similarities, such as shared interest in *replication*. Creating new LDA models for these data also suggested a larger focus on theoretical content (vs. statistics) for the commentaries.

Finally, for another comparison, we review a human-coded analysis of 402 actual reviews from several APA journals (note that the categories are not entirely comparable, and the era is earlier; D. W. Fiske & Fogg, 1990): The most frequent comments concerned interpretations and conclusions (16.1%) and pre-execution conceptualization (15.2%)—statistical analyses (8.5%) and measurement (7.3%) issues were half as frequent. Together with the 1970s crisis' emphasis on theoretical critiques (see opening paragraph), this suggests that blogs focus on methods/statistics more, and conceptualization less, than some other forms of peer review.

Blogs about Whom?

In an initial look at some of the criticisms that have been raised against current social media practices, we explored the frequency of individual researcher mentions. Blogs mentioned at least one of the judge-nominated researchers (viewed as having been discussed critically) in an average of 11% of posts, and specific nominated researchers were mentioned in up to 154 posts in a little over 6 years. Posts with person mentions received more views, social media impact, comments and citations, and these results were generally larger for judge-nominated researcher mentions (vs. non-nominated names).

Compared to posts with no nominated researcher mentions, those with mentions talked less about statistics and fraud but more about replication and research findings. This might suggest that posts mentioning nominated researchers tended to focus on high-level discussions of replication, power, and experimental design—but less on more technical statistical details.

Implications for Controversies, Equity, Methods, and History

This exploration has several potential implications. Given that online social media readers who report feeling personally attacked and seeing others being attacked also report feeling disengaged from the field (West & Skitka, 2017), the kinds of examinations just summarized may aid understanding how to most effectively communicate scientific criticism through blogs. For example, some have suggested that target gender may play a role in online criticism (e.g., Elsesser, 2018). Speaking to this issue, our data show no large discrepancies in the gender of nominated researchers, and in fact male researchers were mentioned by (particularly male) bloggers more often. This does not necessarily speak to the presence or absence of bias, which would require more complex analyses of, for example, base rates (e.g., gender distributions, publication outcomes), but these numbers are a first step in this direction. Future analyses can consider a variety of information and explore details such as the nature of

individual researcher mentions (see Supplement for a limited attempt). Proposed civility norms include focusing on ideas, not the person; promoting inclusivity by attending to power differences; and fostering cooperation, openness, curiosity, and learning (Ledgerwood, 2017).

Our results may also speak to other discussions on how to improve online discussion. Despite blogs being able to provide an open-access platform for diverse voices, we find discrepancies in who uses blogs, and to what effect. For example, the larger number of men with blogs in our sample, despite estimates of a gender-balanced or female-majority psychology during the data production period (e.g., Burrelli, 2008; NSF, 2015), is in line with previous reports (e.g., Weinstein & Sumeracki, 2017) suggesting that men use social media more actively (in this case, by having blogs more often). As with the topic analysis, we looked at the gender of 1039 responses in the journal *Behavioral and Brain Science* using an automated gender classification algorithm (thus estimates are noisy; see Supplement). Results suggested that, in line with the results from the blogs, the majority of authors in this more traditional format of open peer-review were male (63.3% male vs. 31.7% female). However, in our blog data, the women who did have blogs posted more often (based on medians) than men. Nonetheless, male-led blogs had more citations, social media impact, and comments per post, a finding that replicates the pattern in scientific journals (see Eagly & Miller, 2016). Additionally, established researchers had more posts, and their posts had more impact. Thus, incentives and priorities associated with who speaks in social media may be a function of career-stage and demographics.

In general terms, the questions we deal with here have implications for issues such as the consideration of blog posts as legitimate citations and otherwise regarded similarly to traditional peer-reviewed publications, or the role that blogging could play in hiring and promotion decisions. For example, our results suggest that indeed some are already citing blogs in outlets indexed by Google Scholar, and in turn these citation counts may inform career-status decisions.

Our data speak in other ways to these issues, but mostly, we provide an initial step in opening up the use of more exploratory methods to inform discussions about these and other topics in the role of social media in science communication.

Limitations and Future Directions

To be sure, this report has some limitations. For example, the absence of directly applicable base rates or direct comparison groups makes it difficult to extrapolate some of our findings to several current discussions surrounding online forums (e.g., diverse representation). The presented gender base rates might not be the most applicable (e.g., some bloggers were not psychologists), and using bloggers' journal and book citations as a comparison for blog citations is not an ideal anchor. Appropriate comparison groups could range from conference discussions to editorial boards, to peer review. Although we have attempted to review or provide data on some of these comparison groups, many are not amenable to data collection. Certainly, issues discussed here about newer scientific outlets may also appear in more traditional forums. The possibility to conduct this exploration, which can examine properties of the communication and medium, in itself is a strength of open media, such as blogs, as a format of scientific communication. Nonetheless, even if the base rates or comparison groups provided do not cover all the appropriate options, estimating demographics or impact is valuable, and provides a historical record for future examinations of how scientific communication in the field continues to evolve.

Besides these issues, we present correlational data and try to keep our models simple, allowing for many plausible explanations (which can be further explored from the data). Moreover, our findings explore a subset of psychology methods blogs (which we believe represents most such blogs at the time of data collection, independently selected, but might be biased toward salient blogs). Hence, we refrain from making strong inferences beyond describing

the data, simply presenting potential interpretations. But we believe this is a step in the direction of using resources such as text analysis to ground discussions about science communication in psychology. Future research could expand on this approach by incorporating qualitative work, and by drawing additional relevant comparisons with other traditional publication practices. Follow-up work could explore the content and impact of other relevant social media, including Facebook discussion groups and Twitter. Each of these platforms has unique attributes (e.g., Tweets have a word limit of 280 characters vs. the average post length of 1116 words; Facebook groups have multiple posters and sometimes moderators); these could make them prone to specific content and differential reach.

As scientific forms of communication change, tracking these adjustments will help monitor how they impact scientific discussion. In current and future crises in psychology and other fields, understanding the conversation may help us more effectively improve the science.

References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278. doi:10.1016/j.jml.2012.11.001
- 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.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407-425. doi: 10.1037/a0021524
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003) Latent dirichlet allocation. *Journal of machine learning research*, 3, 993-1022. Available at:
<http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Burrelli, J. (2008). Thirty-three years of women in S&E faculty positions (Publication No. NSF 08-308). Arlington, VA: National Science Foundation. Available at:
<https://wayback.archive-it.org/5902/20160210222049/http://www.nsf.gov/statistics/infbrief/nsf08308/nsf08308.pdf>. Accessed on June 25, 2018.
- Cikara, M., Rudman, L., & Fiske, S. (2012). Dearth by a thousand cuts?: Accounting for gender differences in top-ranked publication rates in social psychology. *Journal of Social Issues*, 68(2), 263-285. doi: 10.1111/j.1540-4560.2012.01748.x
- Coan, J., Dunham, Y., Durante, K., Finkel, E., Gabriel, S., Giner-Sorolla, R., Inzlicht, M., Johnson, C., Ledgerwood, A., Spencer, S., Srivastava, S., Van Bavel, J., & Vazire, S. (2017). Promoting open, critical, civil, and inclusive scientific discourse in psychology.

- Available at: <http://www.ipetitions.com/petition/the-tenor-of-discussions>. Accessed June 25, 2018.
- Deutsch, M. (1976). Theorizing in social psychology. *Personality and Social Psychology Bulletin*, 2(2), 134-141. doi: 10.1177/014616727600200214
- Diener E., Oishi S., Park J. (2014). An incomplete list of eminent psychologists of the modern era. *Archives of Scientific Psychology*, 2, 20–32. doi:10.1037/arc0000006
- Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology*, 60(5), 685. doi: 10.1037/0022-3514.60.5.685
- Eagly, A. H., & Miller, D. I. (2016). Scientific Eminence: Where Are the Women? *Perspectives on Psychological Science*, 11(6), 899-904. doi: 10.1177/17456916166663918
- Elsesser, K. (2018). Power Posing Is Back: Amy Cuddy Successfully Refutes Criticism. Forbes.com. Available at <https://www.forbes.com/sites/kimelsesser/2018/04/03/power-posing-is-back-amy-cuddy-successfully-refutes-criticism/>. Accessed June 27, 2018.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, 12(2), 121. doi: 10.1037/1082-989X.12.2.121
- Engber, D. (2017, May). Daryl Bem proved ESP is real: Which means science is broken. *Slate*. Available at: <https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html>. Accessed November 29, 2018.
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116(3), 429-456. doi: 10.1037/0033-2909.116.3.429
- Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. *Proceedings of the 43rd Annual*

Meeting of the Association for Computational Linguistics (ACL 2005), 363-370.

Retrieved from: <http://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf>

Fiske, D. W., & Fogg, L. F. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist*, *45*(5), 591.

Fiske, S. T. (2016). A call to change science's culture of shaming. *APS Observer*, *29*(10), 3.

Greenwald, A. G. (1976). An editorial. *Journal of Personality and Social Psychology*, *33*(1), 1-7.

doi: <http://dx.doi.org/10.1037/h0078635>

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. *The Annenberg Public Policy Center Report Series*, 26.

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

Lakens, D. (2017, April 14). Five reasons blog posts are of higher scientific quality than journal articles. *The 20% Statistician*, Available at <http://daniellakens.blogspot.fr/2017/04/five-reasons-blog-posts-are-of-higher.html>. Accessed June 10, 2017.

Ledgerwood, A. (2017). What do we want our scientific discourse to look like? *APS Observer* *30*(1):18-19.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1-33. 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.
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E., Ware J. J., & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. doi: 10.1038/s41562-016-0021
- National Science Foundation (2015). Science, engineering, and health doctorate holders employed in universities and 4-year colleges, by broad occupation, sex, years since doctorate, and tenure status: 2013. *Women, Minorities, and Persons with Disabilities in Science and Engineering*. Available at <https://www.nsf.gov/statistics/2017/nsf17310/static/data/tab9-24.pdf>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology’s Renaissance. *Annual Review of Psychology*. doi: 10.1146/annurev-psych-122216-011836
- 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.
- PsychBrief (2017b). Psychological Methods Blog Feed. *PsychBrief*, Available at <http://psychbrief.com/psychological-methods-blog-feed/>. Accessed June 10, 2017.
- Research Excellence Framework (2014). FAQ’s. Available at <http://www.ref.ac.uk/about/guidance/faq/all/>. Accessed on June 10, 2017.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641. doi: <http://dx.doi.org/10.1037/0033-2909.86.3.638>
- Sabeti, P. (2018). For better science, call off the revolutionaries. *Boston Globe*, January 21. Available at <https://www.bostonglobe.com/ideas/2018/01/21/for-better-science-call-off-revolutionaries/8FFEmBAPCDW3IWYJwKF31L/story.html>. Accessed on June 25, 2018

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction:

Broadening perspectives from the replication crisis. *Annual Review of Psychology*.

Smith, R. J. (1978). The future of an illusion: American social psychology. *Personality and*

Social Psychology Bulletin, 4(1), 173-176. doi: 10.1146/annurev-psych-122216-011845

<http://dx.doi.org/10.1177/014616727800400137> .

Weinstein, Y., & Sumeracki, M. A. (2017). Are Twitter and Blogs Important Tools for the

Modern Psychological Scientist? *Perspectives in Psychological Science*. doi:

10.1177/1745691617712266

West, T., & Skitka, L., Society for Personality and Social Psychology Convention Committee

(2017). SPSP Climate Survey. Available at [http://www.psych.nyu.edu/westlab/lab-](http://www.psych.nyu.edu/westlab/lab-resources.html)

[resources.html](http://www.psych.nyu.edu/westlab/lab-resources.html). Accessed June 25, 2018.

Footnotes

¹ Diederik Stapel was officially suspended from Tilburg University on misconduct charges in 2011, https://uvtapp.uvt.nl/tsb11/npc.npc.ShowPressReleaseCM?v_id=4082238588785510

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 count outcomes we added the log of the length of the post as an offset, effectively turning the outcome interpretation into the proportion of words in the post that are allocated to the topic, and excluded posts for which the LDA-preprocessed post length was zero, resulting in the removal of 69 posts. For convergence, we used zero-inflated models for social media impact outcomes. We used logistic regressions for dichotomous outcomes (e.g., whether a name was mentioned or not).

³ It has been discussed informally (we couldn't find a source) that within-subject standardizing (vs. simple centering) is problematic. In our data, it was often necessary for models with topic predictors to do this for model convergence. However, an examination of other standardizations tended not to change conclusions; these explorations are included in the available R code.

⁴ Rate ratios are the values by which the outcome is multiplied for a one-unit increase in the predictor (categorical predictors were dummy coded, with baselines being women [gender], ECRs [career-stage], and name not mentioned [for both nominated and non-nominated name mentions]). Thus, to illustrate, a rate ratio of 2 would suggest that men had twice as many commenters in their posts as women, or that every SD increase in fraud topic proportion results in twice as many social media impact indicators. Models may have additional constraints (e.g., covariates, variable coding) that alter interpretations.

⁵ Given that men talked more about many of the topics we focused on, we did a secondary analysis of blog-level aggregated data, not accounting for career-stage, to look at gender

differences across all topics. At this level of analysis, women talked more about content topics such as social science, general psychology, teaching, and theory.

⁶ Only citations for publications after 1970 and with an associated year were included in order to minimize some misclassifications in Google Scholar. For blogs with multiple authors we use an average of all authors with available data.

⁷ For female-led blogs post citations are a poorer proxy for impact, as 90% of all self-citations were by women (particularly one blogger), accounting for 43% of female-led post citations.

Table 1. Topic Proportions

	Mean	SD		Mean	SD
Statistics	0.21	0.17	Clinical	0.06	0.09
Bayesian	0.08	0.09	Neuroscience	0.05	0.05
General	0.04	0.04	Social Science	0.05	0.04
Software	0.04	0.04	Social Science	0.04	0.03
Regression	0.02	0.02	Demographics	0.01	0.01
Visualization	0.02	0.03	Fraud	0.03	0.02
Frequentist	0.01	0.02	Teaching	0.03	0.02
Replication	0.19	0.15	Psychology: Other	0.02	0.02
Science Communication	0.15	0.08	Other Content	0.01	0.01
Research Findings	0.12	0.05	Nutrition	0.01	0.01
Theoretical	0.07	0.05	Law	0.00	0.00
Experimental	0.05	0.05	Unidentified	0.08	0.07

Note. Topics and the mean and standard deviation of their share of words per post (at the blog level, averaged across each blog's posts). Non-bolded topics are subtopics of the bolded topics above them.

Table 2. Topic analyses.

Outcome	Predictor	Rate ratio	2.5% CI	97.5% CI
Statistics	Gender	1.14	0.85	1.53
Statistics	Career-stage	0.71	0.50	1.02
Statistics	Nominated name	0.79	0.73	0.86
Statistics	Non-nominated name	1.12	1.05	1.19
Research findings	Gender	1.22	0.94	1.59
Research findings	Career-stage	0.84	0.65	1.09
Research findings	Nominated name	1.29	1.19	1.40
Research findings	Non-nominated name	1.35	1.26	1.44
Replication	Gender	2.00	1.17	3.41
Replication	Career-stage	0.97	0.61	1.56
Replication	Nominated name	2.00	1.76	2.28
Replication	Non-nominated name	1.38	1.24	1.54
Fraud	Gender	1.80	1.19	2.73
Fraud	Career-stage	1.09	0.71	1.67
Fraud	Nominated name	1.58	1.33	1.88
Fraud	Non-nominated name	0.92	0.79	1.06

Note. Outcomes are the proportion of words per post referring to each topic (more specifically, the number of topic words per post with the log of the number of words in the post as an offset). Due to zero post lengths, 69 posts were excluded from these analyses. Gender and Career-stage predictors control for each other. Predictors baselines are: women, ECRs, and name not mentioned.

Table 3. Comments analyses.

Outcome	Predictor	Rate ratio	2.5% CI	97.5% CI
Comments ^a	Gender	2.28	1.57	3.31
Comments	Career-stage	2.16	1.37	3.4
Comments	Statistics (g)	1.01	0.88	1.15
Comments	Statistics (m)	1.07	0.82	1.39
Comments	Research Findings (g)	1.01	0.91	1.13
Comments	Research Findings (m)	1.15	0.74	1.77
Comments	Replication (g)	1.17	1.05	1.3
Comments	Replication (m)	1	0.78	1.27
Comments	Fraud (g)	0.96	0.88	1.06
Comments	Fraud (m)	0.89	0.75	1.05
Comments	Nominated name	1.96	1.78	2.16
Comments	Non-nominated name	1.54	1.43	1.65
Commenter ^b	Gender	1.99	1.45	2.73
Commenter	Career-stage	1.96	1.33	2.89
Commenter	Statistics (g)	0.98	0.88	1.09
Commenter	Statistics (m)	1.01	0.8	1.28
Commenter	Research Findings (g)	1.1	1.08	1.12
Commenter	Research Findings (m)	1.04	0.75	1.45
Commenter	Replication (g)	1.19	1.17	1.22
Commenter	Replication (m)	0.99	0.8	1.21
Commenter	Fraud (g)	0.97	0.89	1.05
Commenter	Fraud (m)	0.93	0.8	1.07
Commenter	Nominated name	1.64	1.52	1.78
Commenter	Non-nominated name	1.33	1.26	1.42

Note. Outcomes are (a) number of comments per post and (b) number of distinct commenters per post. Gender and Career-stage predictors control for each other. For topic predictors, (g) indicates that post-level standardized predictor (i.e., indicating average differences in topic proportions between posts within a blog), and (m) indicates the blog-level standardized predictor (i.e., indicating differences between the average topic proportions between blogs), controlling for each other. Categorical predictors baselines are: women, ECRs, and name not mentioned.

Table 4. Social media, views, and post citations analyses.

Outcome	Predictor	Rate ratio	2.5% CI	97.5% CI
SM impact ^a	Gender	2.45	1.18	5.07
SM impact	Career-stage	5.77	2.02	16.47
SM impact	Statistics (g)	0.89	0.77	1.04
SM impact	Statistics (m)	0.65	0.43	0.98
SM impact	Research Findings (g)	0.98	0.78	1.23
SM impact	Research Findings (m)	0.82	0.42	1.62
SM impact	Replication (g)	1.41	1.15	1.74
SM impact	Replication (m)	1.32	0.89	1.95
SM impact	Fraud (g)	0.97	0.85	1.09
SM impact	Fraud (m)	1.14	0.84	1.55
SM impact	Nominated name	1.92	1.29	2.85
SM impact	Non-nominated name	3.08	1.86	5.08
Views ^b	Gender	2.72	0.94	7.87
Views	Career-stage	1.35	0.5	3.65
Views	Statistics (g)	1.16	0.95	1.4
Views	Statistics (m)	1.15	0.74	1.77
Views	Research Findings (g)	1.03	0.89	1.19
Views	Research Findings (m)	1.17	0.68	2.03
Views	Replication (g)	1.23	1.14	1.32
Views	Replication (m)	1.41	0.95	2.09
Views	Fraud (g)	0.83	0.77	0.89
Views	Fraud (m)	0.95	0.74	1.21
Views	Nominated name	1.68	1.35	2.1
Views	Non-nominated name	1.54	1.24	1.9
Citations ^c	Gender	2.28	1.07	4.85
Citations	Career-stage	2.52	1.03	6.18
Citations	Statistics (g)	1.11	0.88	1.39
Citations	Statistics (m)	1.08	0.74	1.57
Citations	Research Findings (g)	1.12	1.02	1.24
Citations	Research Findings (m)	0.97	0.63	1.49
Citations	Replication (g)	1.55	1.43	1.68
Citations	Replication (m)	1.25	0.99	1.58
Citations	Fraud (g)	1	0.9	1.11
Citations	Fraud (m)	0.87	0.7	1.07
Citations	Nominated name	2.95	2.22	3.93
Citations	Non-nominated name	1.71	1.23	2.38

Note. Outcomes are (a) number of social media likes/comments/shares per post, (b) number of views per post, and (c) citations per post. Gender and Career-stage predictors control for each other. For topic predictors, (g) indicates that post-level standardized predictor (i.e., indicating average differences in topic proportions between posts within a blog), and (m) indicates the blog-level standardized predictor (i.e., indicating differences between the average topic proportions between blogs), controlling for each other. Categorical predictors baselines are: women, ECRs, and name not mentioned.