SUPPLEMENTAL INFORMATION FOR:

Open Science, Communal Culture, and Women's Participation in the Movement to Improve Science

Mary C. Murphy*, Amanda Mejia*, Jorge Mejia*, Xiaoran Yan*, Sapna Cheryan, Nilanjana Dasgupta, Mesmin Destin, Stephanie A Fryberg, Julie A. Garcia, Elizabeth L. Haines, Judith M. Harackiewicz, Alison Ledgerwood, Corinne A. Moss-Racusin, Lora E Park, Sylvia, P., Perry, Kate A. Ratliff, Aneeta Rattan, Diana T. Sanchez, Krishna Savani, Denise Sekaquaptewa, Jessi L. Smith, Valerie Jones Taylor, Dustin B. Thoman, Daryl A. Wout, Patricia L. Mabry**, Susanne Ressl**, Amanda Diekman**, Franco Pestilli**

* shared first author contribution
** shared senior author contribution

ABSTRACT

Science is undergoing rapid change with the movement to improve science focused largely on reproducibility/replicability and open science practices. This moment of change—in which science turns inward to examine its methods and practices—provides an opportunity to address its historic lack of diversity and non-inclusive culture. Through network modeling and semantic analysis, we provide an initial exploration of the structure, cultural frames, and women’s participation in the open science and reproducibility literatures (N = 2,926 articles and conference proceedings). Network analyses suggest that the open science and reproducibility literatures are emerging relatively independently of each other, sharing few common papers or authors. We next examine whether the literatures differentially incorporate collaborative, prosocial ideals that are known to engage members of underrepresented groups more than independent, winner-takes-all approaches. We find that open science has a more connected, collaborative structure than does reproducibility. Semantic analyses of paper abstracts reveal that these literatures have adopted different cultural frames: open science includes more explicitly communal and prosocial language than does reproducibility. Finally, consistent with literature suggesting the diversity benefits of communal and prosocial purposes, we find that women publish more frequently in high-status author positions (first or last) within open science (vs. reproducibility). And, this finding is further patterned by team size and time. Women are more represented in larger teams within reproducibility. And, women’s participation is increasing in open science over time, and decreasing in reproducibility. We conclude with actionable suggestions for cultivating a more prosocial and diverse culture of science.

Author Contributions. Literature search: MCM, ABD, PLM, FP, CRS, EH, AL, KR, ND, DT, LP, JS, SP, JH, KS, VJT, AR. SC. SF, CAM-R, DW. AL, DS, JAG, DTS. Figures: XY, AM, PLM, SR Study design: MCM, PLM, ABD, SR, FP, AM, JMM. Data Collection: XY, ELH, AL, KAR, JMM. Data analyses: AM, XY, JMM, PLM. Data interpretation: AM, XY, JMM, PLM, MCM, FP. Writing: MCM, FP, SR, PLM, ABD, AM, CRS, ELH, AL, KAR, ND, DT, LP, JLS, SPP, JMH, KS, VJT, AR, SF, MD, AR, SAF. Effectively three groups of authors contributed to this work with different degrees of contribution. Author Block 1: First author contributors: MCM, AM, JMM, XY. Author Block 2: Senior contributors: PLM, SR, AD, and FP. Author Block 2: Study supporters.

Acknowledgements. M.C.M. was supported by NSF CAREER DRL-1450755 and NSF HRD-1661004 and by a Russell Sage Foundation grant (87-15-02). A.B.D. was supported by NSF GSE-1232364. F.P. NSF OAC-1916518, NSF IIS-1912270, NSF IIS-1636893, NSF BCS-1734853, NIH 1R01EB029272-01, Google Cloud Research Award, a Microsoft Investigator Fellowship, the Indiana University Areas of Emergent Research initiative “Learning: Brains, Machines, Children,” the Indiana University Pervasive Technology Institute. The authors would like to acknowledge Angela Sharpe and Cassidy Sugimoto for their thoughtful discussion about the topics in the paper which were reflected in the manuscript, and Michael Jackson for help with the design of the figures.
Data Collection & Analysis

A total of 11,338 original papers were collected using the snapshot of Microsoft Academic Graph (MAG; https://academic.microsoft.com; (1) on Feb 23, 2018. To collect the datasets, we searched MAG for all publications with specific “field of study tags” as “open science” or “reproducibility”. Among all the records, only 68 papers were categorized as both “open science” and “reproducibility”. Moreover, of the 36,296 unique author ids represented in these literatures, very few (N = 457) have authored in both literature. These findings suggest that the two literatures are developing rather independently. For the purposes of our analyses, we removed papers that were categorized as both “open science” and “reproducibility” to avoid double-counting papers and skewing analyses. Among the remaining records, we only considered formal published papers of the type “journal” or “conference”. The resulting dataset includes 3,431 Open Science papers and 7,839 Reproducibility papers.

Among the remaining records, we only considered formal published papers of the type “journal” or “conference” (document types “book”, “book chapter” and “patent” were removed). We also removed 43 papers with duplicate titles. We examined the remaining number of papers being published in each year within each literature, shown in the left panel of Figure S1. As very few Open Science papers were published prior to 2010, and few papers in either field have been published in 2018, in the following analyses we only use data for papers published between 2010 and 2017, which includes 879 Open Science papers and 2,047 Reproducibility papers.

For each author, we use the gender package in R (2); R package version 0.5.0.9000 (https://github.com/ropensci/gender) to predict the probability of each name being “female”. We define name parts as being separated by a space and consider all but the last part of each name (representing the last name) for gender detection. We exclude parts representing initials and apply gender detection to the first remaining part, resulting in a probability of being a woman author with at least one usable name part. Authors with probability over 0.5 are labeled “female” and those with probability below 0.5 are labeled “male”. Authors with no usable name parts are labeled “unknown”.

Figure S1. Distribution of year of publication and team size for papers in the Open Science and Reproducibility literatures. Top panel: The open science literature has developed primarily since 2010. We therefore limit all of our analyses to papers published after 2010. Bottom panel: Very few papers in either discipline involve more than 12 authors. We therefore exclude authors with more than 12 papers from regression models, since they may have high influence on the model fit and represent a small minority.

Conceptual validation of our paper categorization method: Exploring the categorization of “computational reproducibility” papers.

A thoughtful reviewer suggested that given the various possible meanings of the term “reproducibility” in different scientific fields (see (3) for a detailed discussion of this issue), we explore how papers on “computational reproducibility” were categorized by our process. As a conceptual check on our paper categorization approach, we empirically explored how our
method categorized various “computational reproducibility” papers with different MAG field of study tags. Specifically, we conducted an exploratory analysis that empirically examined where “computational reproducibility” papers ended up—in “open science” or in “reproducibility.”

One challenge we immediately faced is that there is not a clear operationalization of “computational reproducibility”. However, we relied heavily on Victoria Stodden’s work (4, 5) in choosing the field of study tags that helped us identify computational reproducibility papers for the purposes of this analysis. Specifically, we used the MAG fields of study tags: “workflow”, “repeatability”, "data sharing", "open data", and "software" as proxies for "computational reproducibility". Table S1 summarizes the number of papers associated with each field of study and how our original categorization process grouped those papers:

|                | Workflow | Repeatability | Data sharing | Open data | Software |
|----------------|----------|---------------|--------------|-----------|----------|
| Categorized as Open Science | 39       | 0             | 76           | 267       | 103      |
| Categorized as Reproducibility | 33       | 260           | 11           | 2         | 79       |

Table S1. Exploring the categorization of computational reproducibility papers. All papers with the MAG field of study tag “repeatability” were categorized by our method as Reproducibility papers, as intended. And, almost all papers with the MAG field of study tag “open data” or “data sharing” were categorized as “open science” papers, as intended.

As can be seen in the highlighted columns, our categorization method sorted papers as intended. All papers that have the field of study tag “Repeatability” were categorized by our method as “reproducibility” papers—in line with the NAS’s conceptualization of reproducibility (3). And, almost all papers that have the MAG field of study tag “Open data” were categorized by our method as “open science” papers. Moreover, almost all papers with the MAG field of study tag “Data Sharing” were also categorized as “open science” papers, as we intended. Because the remaining computational reproducibility papers with “Workflow” and “Software” field of study tags are relatively evenly spread across our “open science” and “reproducibility” categories, they do not bias our findings; that is, how these papers are categorized does not materially affect our results. These results underscore the validity of our conceptual categories and bolster our confidence in our categorization method.

Network Analysis

Based on the sample between 2010 and 2017, we constructed the paper co-authorship networks for 879 Open Science papers and 2,047 Reproducibility papers. Each node represents a scientific article. Two nodes share an edge if at least one author appears in both
papers. Based on exact string matching of author names, we identified 3,157 unique author names in the open science literature and 8,766 in the reproducibility literature. In the open science literature, the network contains 389 edges (i.e., pairs of papers with at least one author in common), and 856 edges in the reproducibility literature. The network visualizations are generated using Gephi (6).

**Robustness check: Community structure differences in open science and reproducibility—for papers with multiple authors.**

In addition to the two collaboration networks in the main text, we examined the robustness of our findings when examining only multi-authored papers (excluding single-authored papers). While the original analysis effectively visualizes greater connection in open science and greater fragmentation in reproducibility due to the prevalence of single-authored papers, it is also possible that some edges in the original networks could reflect a few single-authored papers authored by the same individual—effectively “connecting” the individual to themselves.

Are our results robust to an analysis limited to only multi-authored papers? In this new analysis, single-authored papers are removed. In Open Science, this meant removing 255 papers (nodes) and 95 edges and in Reproducibility this meant removing 340 papers and 209 edges. Nodes now represent scientific articles that have at least two authors; two nodes share an edge if at least one author appears in both papers. Results revealed that this Open Science network contained 624 nodes and 294 edges, while the Reproducibility network contained 1,707 nodes and 647 edges. Importantly, the Open Science network remains more edge dense (0.151%) than the Reproducibility network (0.044%), demonstrating again a higher degree of collaboration in the Open Science literature (one-sided Fisher’s exact test p < 0.001) even among multi-authored papers. The “connected components analysis” also confirms that the Reproducibility network contains more isolated subgroups than the Open Science network with 1,364 versus 452 components, respectively. The size differences of the two networks are also consistent. We find that the average component size (ACS) remains higher for the Open Science network (ACS: 1.383 vs. 1.25).
Figure S2. Differences in author community structure limited to multi-authored papers: Open Science vs. Reproducibility. Visualization of the two networks to facilitate interpretation of the observed differences between the two literatures in terms of network connectedness and fragmentation.

Semantic Text Analysis of Scientific Articles’ Abstracts

Below, we provide the list of terms used to measure prosocial and communal values in the Open Science and Reproducibility literatures sourced from (7).

| accepting | confide | encourag | honest | nonjudgmental | sensitiv | tutor |
|-----------|---------|----------|--------|---------------|----------|-------|
| accommodat| conscient| environment| honorable| nonprofit | serv | underst |
| affect | conservation | equal | honourable | not-for-profit | share | universal |
| agreeable | considerate | ethic | hospital | nurtur | shari | unprejudiced |
| aid | contribut | everybody | human | peace | shield | upright |
| altruis | cooperat | everyone | impartial | philanthrop | sincer | virtuous |
| appreciat | cope | facilitat | inspiring | prais | societ | volunteer |
| approachable | coping | fair | integrat | prejud | solidarit |
| assist | courteous | forgiv | integrity | protect | support |
| benefit | courtesy | freed | interact | reciproc | sustainab |
| benevolen | defend | genero | invit | relia | sympath |
| biodivers | dependab | gentle | involv | relied | taught |
| care | dignity | genuin | justice | rely | teach |
| caring | donat | giv | kids | respectful | team |
Women’s Participation in High-Status Authorship Positions

Figure 3 in the main text displays the gender mix in high-status authorship positions in both literatures. For single-author papers, we observe that more papers are written by men in both literatures, based on an exact binomial test for the hypothesis that the proportion of women is less than 0.5 ($p < 0.0001$ for both literatures). For this analysis, we exclude any papers with non-observable author gender. We do not observe a statistically significant difference in rates of women’s representation in high-status authorship positions in single-author papers across the two fields—although women’s representation falls short of gender parity in both literatures. We next focus on the multiple-author case.

For multiple-author papers, we observe somewhat higher rates of women in high-status positions in the Open Science literature compared to the Reproducibility literature. Descriptively, women hold a high-status authorship position in at least 50% of the multiple-author papers in the open science literature, while that number is 35% in the reproducibility literature. These numbers are likely conservative estimates, since there could be additional representation of women in high-status authorship positions among the papers with unknown author genders. Excluding papers where both high-status positions have unknown gender, or one is male and the other is unknown, the rates of women in high-status authorship positions are 63% in open science and 58% in reproducibility. We next use regression analysis to formally compare the rates of women in high-status positions in the two literatures while controlling for team size, year of publication, and publication type.

Regression Analysis

Women are more likely to be represented in high-status author positions in Open Science (vs. Reproducibility)—main text analysis details. Table S3 reports estimates and confidence intervals on the odds scale (where odds of female lead authorship is defined as $Pr(Y_i = 1)/(1 - Pr(Y_i = 1))$) for each parametric (i.e., non-spline) coefficient in our main regression model. The estimates and intervals represent multiplicative effects on the odds of having a female in a high-status position. A value greater than 1 therefore represents a positive
effect, while a value below 1 represents a negative effect, and a value of 1 represents no effect. Odds $x$ can be transformed to a probability $p$ using the transformation $p = x/(1 + x)$.

|                                    | Estimate | 95% Confidence Interval |
|------------------------------------|----------|-------------------------|
| Intercept                          | 2.785    | 2.029                   | 3.822 |
| Reproducibility                   | 0.393    | 0.259                   | 0.597 |
| Publication Year                   | 1.156    | 1.045                   | 1.278 |
| ConferencePaper                   | 0.681    | 0.477                   | 0.971 |
| Reproducibility X Publication Year | 0.834    | 0.743                   | 0.936 |
| Reproducibility X Team Size (Spline) | n/a      | n/a                     | n/a  |

Table S3. Coefficient estimates and 95% confidence intervals for the parametric coefficients of the multi-author logistic regression model. Coefficient values are exponentiated to show multiplicative effects of each variable on the odds of having a female in a high-status position. Normal confidence intervals are computed on the log-odds scale then exponentiated to obtain the interval on the odds scale. Intervals containing only values greater than 1 indicate significant positive effects, shown in bold black font; intervals containing only values less than 1 indicate significant negative effects, shown in bold red font. Team size effects are modeled using flexible splines and therefore do not have coefficient estimates or confidence intervals, shown in gray font.

Robustness check: Women’s representation in high-status author positions in Open Science (vs. Reproducibility)—Controlling for field of study.

We performed several robustness checks to control for differences across academic fields of study. Specifically, we fit two alternative model specifications—one that controlled for field of study and a second that stratified by field of study (see the following section for stratification analyses and results). First, we refit the model including indicators for academic fields of study as control variables. Because each paper lists multiple fields and there are over 2,000 unique fields listed, we considered the 30 most frequently appearing fields, which constitute over 44% of all field appearances. Of these 30, we removed several fields not considered traditional fields of study (“publishing”, ”workflow”, ”data sharing”, ”reproducibility”, ”open science”, ”repeatability”, ”open data”, ”data mining”, ”intraclass correlation”) and combined several medicine-related fields into a single “medicine” field (“anesthesiology”, ”cardiology”, ”diabetes mellitus”, ”internal medicine”, ”surgery”, ”alternative medicine”, ”physical therapy”, ”pathology”, ”radiology”, ”medicine”). This resulted in 12 academic fields of study (“medicine”, ”artificial intelligence”, ”management science”, ”analytical chemistry”, ”bioinformatics”, ”engineering”, ”knowledge management”, ”psychology”, ”software”, ”biology”, ”statistics”, ”computer science”). For each field, we constructed an indicator variable equaling 1 if the field was listed by each paper among its fields of study and 0 otherwise. We controlled for field of study by including these indicator variables as binary covariates in the model specified above. Of the 1,409 papers included in the original model, only 98 did not list any of these fields and were excluded from this model.
Table S4 (below) reports estimates and 95% confidence intervals on the odds scale (where odds of female lead authorship is defined as \( \frac{Pr(Y_i = 1)}{1 - Pr(Y_i = 1)} \)) for each parametric (i.e., non-spline) coefficient in our regression model controlling for the 12 most common academic fields of study. The estimates and intervals represent multiplicative effects on the odds of having a woman in a high-status position. A value greater than 1 therefore represents a positive effect, while a value below 1 represents a negative effect, and a value of 1 represents no effect. Odds \( x \) can be transformed to a probability \( p \) using the transformation \( p = \frac{x}{1 + x} \). In short, the effect of belonging to the Reproducibility literature remains negative and statistically significant, and the trend over time is again positive within the Open Science literature and slightly negative within the Reproducibility literature.

| Coefficient                           | Estimate | 95% Confidence Interval |
|---------------------------------------|----------|-------------------------|
| Intercept                             | 2.412    | 1.472 - 3.951           |
| Reproducibility                       | 0.489    | 0.311 - 0.768           |
| Publication Year                      | 1.124    | 1.004 - 1.257           |
| Conference Paper                     | 0.801    | 0.531 - 1.209           |
| Team Size (Spline; no estimate generated) | n/a      | n/a - n/a              |
| Reproducibility X Publication Year    | 0.856    | 0.754 - 0.973           |
| Reproducibility X Team Size (Spline)  | n/a      | n/a - n/a              |
| Medicine                              | 1.058    | 0.741 - 1.510           |
| Artificial intelligence               | 0.983    | 0.521 - 1.856           |
| Management science                    | 1.128    | 0.626 - 2.034           |
| Analytical chemistry                  | 0.454    | 0.255 - 0.810           |
| Bioinformatics                        | 0.566    | 0.339 - 0.947           |
| Engineering                           | 0.953    | 0.647 - 1.404           |
| Knowledge management                  | 1.622    | 0.954 - 2.758           |
| Psychology                            | 0.894    | 0.559 - 1.430           |
| Software                              | 0.907    | 0.585 - 1.407           |
| Biology                               | 1.379    | 0.880 - 2.163           |
| Statistics                            | 1.431    | 0.947 - 2.162           |
| Computer science                      | 0.793    | 0.539 - 1.167           |

Table S4. Coefficient estimates for the parametric coefficients of the multi-author logistic regression model controlling for academic field of study. Coefficient values are exponentiated to show multiplicative effects of each variable on the odds of having a female in a high-status position. Normal confidence intervals are computed on the log-odds scale then exponentiated to obtain the interval on the odds scale. Intervals containing only values greater than 1 indicate significant positive effects, shown in bold black font; intervals containing only values less than 1 indicate significant negative effects, shown in bold red font. Team size effects are modeled using flexible splines and therefore do not have coefficient estimates or confidence intervals, shown in gray font.
Robustness check: Women’s representation in high-status author positions in Open Science (vs. Reproducibility)—Stratifying by field of study.

As a second robustness check, we stratified the model by academic field of study to avoid possible confounding that could be induced by differences in team size, time trends or other covariates across fields. First, we further reduced the number of fields in order to avoid loss of power due to small sample size in the field-specific models. Specifically, we combined into a single field “computer science”, “artificial intelligence” and “software” (“computer science”); “management science” and “knowledge management” (“management”); “biology” and “bioinformatics” (“bioscience”). This resulted in 8 fields of study. For two of these, there were virtually no Open Science papers (“statistics” ($n_{OS} = 3$) and “analytical chemistry” ($n_{OS} = 0$)); therefore, in the models for these fields there are no main or interaction effects for $Rep_i$, and their model coefficients represent only Reproducibility papers. There was also small sample size for Open Science papers in “psychology” ($n_{OS} = 18$) and for Reproducibility papers in “management” ($n_{RR} = 23$); we left the main and interaction effects for $Rep_i$ in these models, but the effects related to the small sample size should be interpreted with caution. All other field-specific sample sizes for Open Science and Reproducibility papers were over 50 (range: 52-180).

Estimates and Visualizations of Team Size and Longitudinal Analyses Stratified by Field of Study.

Figures S3 and S4 display the estimates and 95% confidence intervals for the probability of having a woman in a high-status position as year and team size vary, respectively. Values are logistic regression estimates shown on the probability scale, with 95% confidence interval bands, based on models stratified by the top 8 fields of study. They were constructed in the same way as the left and right panel of Figure 5 in the main text, respectively, but using the model within each academic field of study.
Figure S3. Women’s participation and team size, stratified by the eight most common fields of study. Within nearly every field, we see results consistent with the pattern reported in the main text: women have higher rates of high-status authorship in larger teams within Reproducibility, while rates are comparatively and consistently high in Open Science across team sizes. One notable exception is in Medicine, where rates of women’s participation are similar across Open Science and Reproducibility, and no team size effect is apparent. In Management and Psychology, small sample size in either the Open Science or Reproducibility literatures (indicated with dashed lines) make comparisons between the two literatures difficult within these fields. In Analytical Chemistry and Statistics, few-to-no Open Science papers existed, but the Reproducibility papers showed a similar trend of lower female representation in smaller teams. Values are logistic regression estimates shown on the probability scale, with 95% confidence interval bands. To produce the estimates, the x-axis variable and literature category are varied, while the remaining model variables are fixed (see Methods for details).
Figure S4. Participation of women in high-status authorship positions over time, stratified by the eight most common fields of study. Consistent with the results presented in the main text, we find that, within most fields of study, the representation of women in high-status positions has grown over time in Open Science, while it has declined in Reproducibility. One exception is in Psychology, where women’s participation in Reproducibility has also grown over time. This could be due to several factors including greater number of women scholars who comprise academic Psychology (8) as well as concerted efforts to include women and other underrepresented groups in the movement to improve Psychological Science (as described in the main text discussion). In Management and Psychology, small sample size in either the Open Science or Reproducibility literatures (indicated with dashed lines) make comparisons between the two literatures difficult within these fields. In Analytical Chemistry and Statistics, few-to-no Open Science papers existed, but the Reproducibility papers showed a similar trend of decreased female participation over time. Values are logistic regression estimates shown on the probability scale, with 95% confidence interval bands. To produce the estimates, the x-axis variable and literature category are varied, while the remaining model variables are fixed (see Methods for details).
Robustness check: Semantic text analysis—Stratified by field.

The semantic text analysis described in the main text was repeated; however, these analyses were stratified by the most common fields of study. Starting with the 8 fields identified as above, we excluded “statistics” and “analytical chemistry” due to the lack of Open Science papers in both fields. Figure S5 displays histograms (on the density scale for comparability across literatures) of prosocial word density score within Open Science and Reproducibility papers. A two-sided permutation test for differences in the mean and median scores within each field shows that the Open Science literature includes significantly more frequent use of communal and prosocial words than does the Reproducibility literature in nearly every field (p-values are reported in Figure S5).

Figure S5. Distribution of communal and pro-social word density of abstracts in the Open Science and Reproducibility literatures, stratified by the eight most common fields of study. Consistent with the results reported in the main text, we see that within almost every field of study, abstracts in the Open Science literature include more words associated with communality and prosociality than those in the Reproducibility literature. The differences in the mean and median are statistically significant in each field with the exception of Psychology. Statistics and Analytical Chemistry are excluded due to the lack of Open Science papers in both fields.
References

1. A. Sinha, et al., An Overview of Microsoft Academic Service (MAS) and Applications in Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion., (ACM, 2015), pp. 243–246.

2. L. Mullen, C. Blevins, B. Schmidt, Gender: predict gender from names using historical data. R package version 0. 5 1 (2015).

3. H. V. Fineberg, et al., Reproducibility and Replicability in Science. The national academies of sciences, engineering, and medicine (2019) (April 1, 2020).

4. V. Stodden, Computational Reproducibility (2018).

5. V. Stodden, J. Borwein, D. H. Bailey, Setting the default to reproducible. computational science research. SIAM News 46, 4–6 (2013).

6. M. Bastian, S. Heymann, M. Jacomy, Others, Gephi: an open source software for exploring and manipulating networks. lcssm 8, 361–362 (2009).

7. J. A. Frimer, K. Aquino, J. E. Gebauer, L. L. Zhu, H. Oakes, A decline in prosocial language helps explain public disapproval of the US Congress. Proc. Natl. Acad. Sci. U. S. A. 112, 6591–6594 (2015).

8. NSF Report: Women, Underrepresented Minorities Gain Ground in Behavioral Science. APS Obs. 33 (2020).