Project report to the American Philosophical Association
“Gender in the Stanford Encyclopedia of Philosophy”
P.I. Colin Allen colin.allen@pitt.edu
September 30, 2017

This project, funded by a small grant from the APA, sought to examine the gender distribution in one of the most influential online resources of philosophy, the Stanford Encyclopedia of Philosophy, by providing statistical measures of female representation in authorship and rates of citation, investigating patterns of citation for potential biases, comparing the SEP gender distribution to the more general demographics of the field, and exploring whether the content of SEP articles written by women is different than by men.

It is important to note at the outset that the gender-binary framing of this project is less than ideal. However, we believe that the kind of data described below are valuable in themselves and provide the basis for a more comprehensive study that would require deeper research into the life histories of the individuals represented in the database. We believe that such data provide a good starting point for that deeper research. We are also mindful of the problem that a significant proportion of names are gender-ambiguous: Francis, Hillary, Jean (français ou anglaise?!), etc. For this reason we did not algorithmically match names to genders, but invested considerable person-hours in tracking down enough biographical information as was needed for the categorization.

Project Goals

Our primary goals were as follows:

1. Generate statistical information quantifying the gender distribution in the SEP. Such statistics would include female authorship rates within subfields of philosophy and rate of women cited within subfields.
2. Identify specific ways in which a gender gap might be perpetuated by philosophers' citation patterns. For example, this project will seek to find whether men are less likely to cite women than women are to do so.
3. Compare the gender distribution of SEP authorship and citations with the gender distribution of academic positions in philosophy.
4. Explore whether differences between the writing of women and men are detectable for a given subfield of philosophy. This will be tested by running articles from each subfield corresponding SEP subject areas through a topic modeler and correlating with gender.

Project Outcomes

Each of the project goals is addressed in detail below, but first an overview: We did not spend all the money awarded, and the unspent portion will be returned to the APA by Indiana University (IU). We were granted a spending extension to the end of September, 2017, so a full accounting of the expenditures will follow once the books have been reconciled by the
corresponding grant officer at IU. We have not yet fully accomplished the goals for a variety of reasons -- partly technical (due to the scale of the task) and partly personnel-related (illnesses and relocations) -- but we have preliminary results in all four areas, and we will finish the project regardless of further funding, as the P.I. has the resources in his new position at the University of Pittsburgh to complete the project going forward.

Details about each of the goals are as follows.

1. We have partially completed the task of generating statistical information quantifying the gender distribution in the SEP. This task had two major components, first, to import the full bibliographies of the SEP and second, to identify cited individuals. For the first subtask, so far we have manually processed 40% of the entries in the SEP. The import of the remaining bibliographic information is a task that can be automated, which will greatly increase the ingest speed, but requires some additional programming work that will be completed by the end of 2017. For the second subtask we have manually identified well over 98% of the authors cited, but there remain challenges. In particular, professional practice in much of the 20th Century was for authors to publish under their initials only. We found this to be particularly challenging for entries in the philosophy of science because this area contains a lot of citations of scientific papers from the early part of the 20th C., and these cited papers have a large number of co-authors who are often difficult to identify from their names and initials alone. Nevertheless, we have managed to track down all but a handful of the most difficult cases.

We have prototyped a way of graphing these results for display on a website, as shown in Fig 1. In addition to completing the data ingestion, and thus providing more complete results, future enhancements to this site will allow users to switch between absolute citation counts, as shown in Fig. 1 and normalized data (percentages of total citations in that area). The data represented in this way thus will permit more detailed analysis of citations to women cited within various subfields of philosophy.

![Graph showing gender citation patterns by subject area.](image)

**Figure 1:** SEP citation patterns by gender and subject area. Please note that these data are not final, and proportions are likely to change once all data have been ingested. Nevertheless, we expect that the Feminist Philosophy subject area will remain as the only SEP subject area to show more citation of female authors than of male authors.
In addition to merely displaying the summary data, we have prototyped functionality that was requested during a discussion at the PSA women’s caucus last fall -- namely to be able to pull up a list of articles in a given subject area by female authors as a tool that will make such work more easily discoverable for scholarly and pedagogical purposes. Fig. 2 shows this in action, with a list of female authored or co-authored articles cited in one of the SEP’s subject areas. We have the underlying software written that will make these citations available for import into various citation management tools, as well as linkable to PhilPapers, and we aim to have that fully implemented by summer 2018.

![Figure 2](image.png)

**Figure 2.** Web interface showing some of the articles by female authors or coauthors cited by entries in the Normative Ethics subject area of the SEP.

2 and 3. The database we have built enables us to quantify rates of citation of female authors according to the gender of the SEP article authors, and to compare such rates with profession-wide distributions of women in academic positions in philosophy. However, given the
partial nature of our current dataset, it would be inappropriate to provide numbers that could lead to incorrect conclusions. We will publish this information once it is complete enough.

4. In the past year we have made our topic modeling of the SEP a regular, ongoing process subject to automatic updates as the encyclopedia evolves (http://inphodata.cogs.indiana.edu/sep/). New jobs for the P.I. and the doctoral student primarily responsible for the topic modeling have delayed implementation of the topics x gender x subfield analysis, but the PI now has a student programmer in Pittsburgh who will be able to complete this task within the next 6 months.

Additional outcomes

We have built a working prototype of a public site for access to data. Figure 3 shows the front page, and Figure 4 shows an example of the kinds of graphs available through this site, in addition to those shown in the figures above.

![Figure 3. Project home page.](image)
Figure 4. Authorship of works cited in the SEP by year of publication of the cited work. These data represent only a subsample of the full works cited within the encyclopedia.

What next?

The lack of full completion of these project goals is not fully excusable, and full responsibility falls on the P.I.’s shoulders for not having kept things moving faster. Weekly meetings with the student workers during the first months of 2017 identified a number of bottlenecks and enabled us to improve the workflow, but some illness and eventual surgery for our most productive worker pushed much of the data collection into July, and delayed recognition of other parts of the project that were not being completed as fast as necessary. Nevertheless, we have constructed all the basic components of the system, and it will be completed at no further cost to the APA during what is currently a sabbatical year for the P.I.