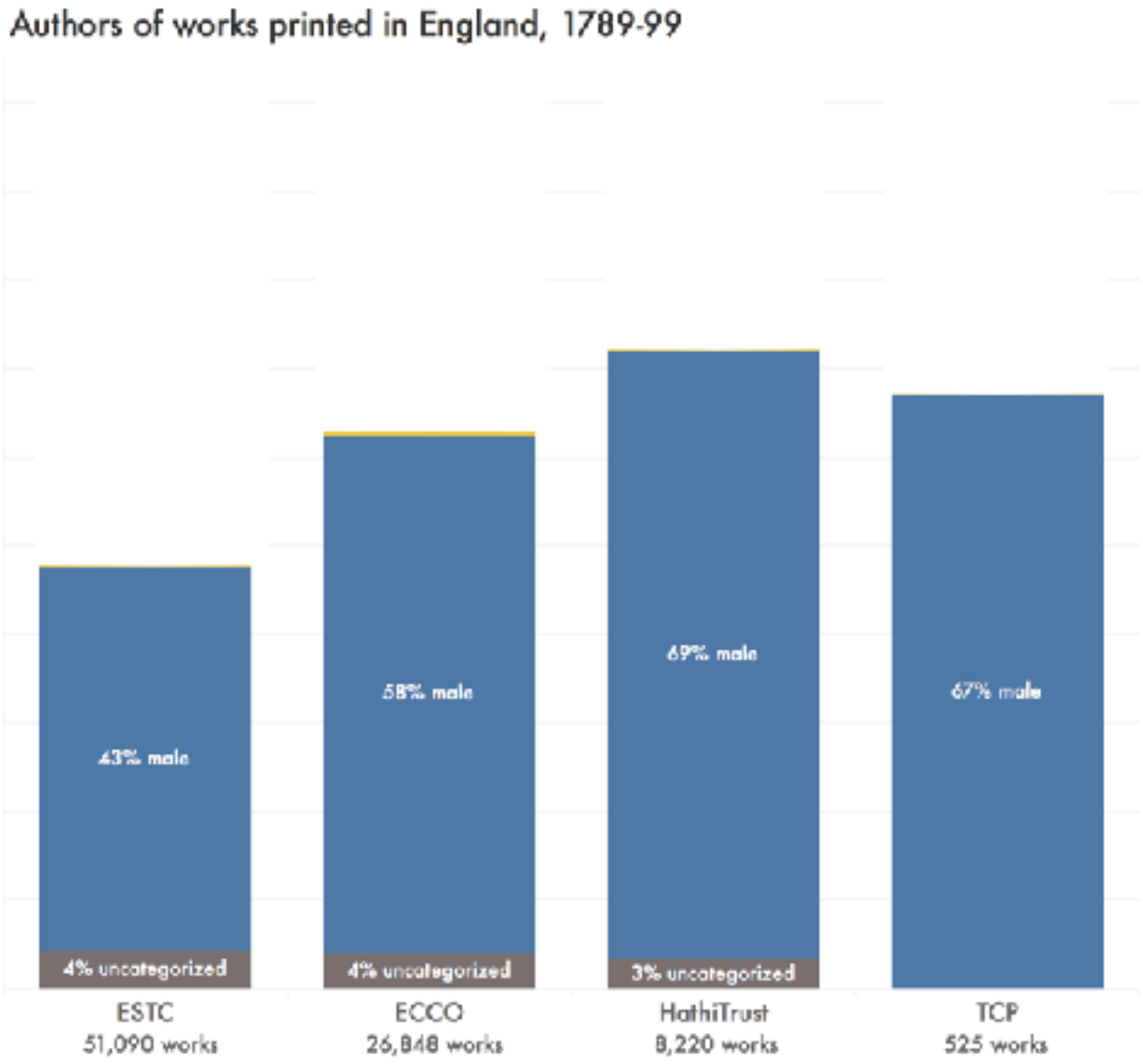
Print Politics in the Digital Archive, 1789-99

Chapter Three

The naive hypothesis, investigating a historical phenomenon, is that factors like systemic sexism are likely to have an effect. As I have shown in chapter two, the digital archives currently used by eigtheenth-century scholars were created during their own historic moments, which influenced their priorities as texts were digitized. Resouces which provide more information about particular works — such as ECCO, which provides PDF scans of microfiche, or the TCP, which provides detailed XML-encoded transcripts — inevitably provide their more thorough information about fewer texts. The ESTC is able to include so many works in its database in part because it includes so little *about* each work: little more than the information found on the title page, and a list of libraries where the original can be consulted. As the more resource-intensive archives created their digital surrogates, therefore, the texts they include are likely to be less indiscriminate, and more strongly influenced by a personal assessment of what works are most important. In many other areas, writing by women has been dismissed as less important or worthy of serious study than writing by men. If we compare the holdings of major digital archives, do the smaller and more resource-intensive archives show evidence of systematically opting not to invest in women's writing?

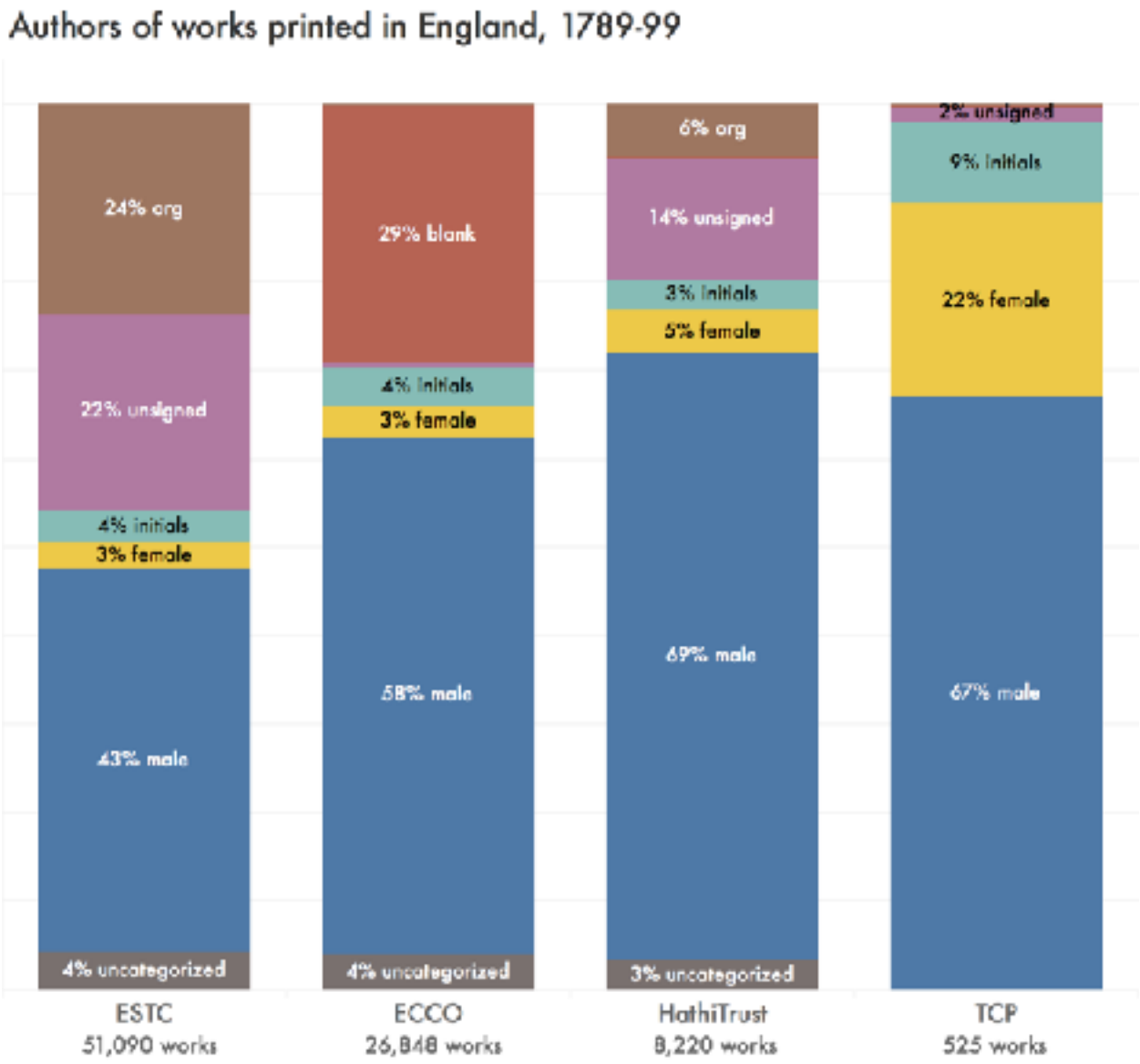
To answer this question, I acquired the full list of all titles included in the ESTC, ECCO, the TCP, and in HathiTrust which were printed in England between 1789 and 1799. In most cases, I was only able to download a superset of the required texts, and then manually removed titles which fell outside of either my geographic or chronological limits. Using OpenRefine, I then manually classified the author's gender for at least 96% of all titles in each database. Most eighteenth century names carry much stronger gendered signalling than contemporary names like "Alex" or "Logan." "John," "Thomas," and "William" are unambiguously male names in the eighteenth century; "Charlotte," "Anne," and "Maria" are equally unambiguous as female names. As I encountered less common names I confirmed the specific identity of any figure with whom I wasn't familiar, as detailed in the appendices. At first glance, the increasing

presence of male authors in smaller and more specialized databases seems to confirm the naive hypothesis about the underrepresentation of women:



The category "male" here includes both named authors and pseudonymous authors whose pseudonyms give clear signals as to their gender. "A Country Clergyman" or "A true lover of his country," for example, may not be identifiable individuals, but they at least claim to represent male authors. This category does not include authors who published only under their initials, since the initials do not give clear evidence of gender, although it seems that most of these initials were understood to represent

specific known men (e.g., W. H. Ireland, who in no way concealed his indentity as William Henry Ireland.) As the archives get smaller, the percentage of works by men increases. However, the rest of the chart is not explained well by historic sexism:



Here we can see that the overlooked group, unable to compete with male authors for archival attention, is not "female authors" but "authors which are not associated with an individual personality." The category "unsigned" here captures works to which no name at all was signed (essentially, anonymous

works, though in this period the absence of a name does not imply the *concealment* of a name as it would in the contemporary definition of "anonymous") as well as pseudonymous works which offered no clues to an author's gender (e.g. "A friend of peace"). "Organizations" include missives from various branches of the government, reports from scholarly and charitable organizations, catalogues from various companies, and other works which are easily rejected as "non-literary." The identifier "blank," for ECCO, appears to capture both unsigned works and works by organizations. ECCO so thoroughly rejects the idea that "Society for the Improvement of Naval Architecture" or "Benevolent Institution for the Sole Purpose of Delivering Poor Married Women at Their Own Habitations" could be considered as "authors" that it does not record them in the "author" field. [Future work will see if I can get this information from another more complicated way of accessing the data, to see whether ECCO duplicates Hathi's preference for unsigned over group-authored works.]

The underrepresentation of unsigned authors cannot be mapped onto the multicultural rhetoric of political representation, the way Guillory has described with gendered representation. "Unsigned" or "pseudonymous" people are not an identity category in the real world requiring equitable social standing: the unsigned and pseudonymous authors, if identified, would all turn out to have names, genders, races, class standings, and all the other traits of humans in the world. What their status as "unsigned" in the database really means is not that those traits are absent — it means that those traits are undetermined at the site of the title page.

Scholars have worked to infrastructurally eliminate the "unsigned" author. Names are placed into the relevant database fields wherever possible, supplying new information that cannot be found on the title page and may have been entirely unavailable to eighteenth century readers. Jane Austen's works, for example, were published unsigned, and then "by a lady," but will always be present in the database under "Austen, Jane."[1] There is obvious practical use in disseminating author identifications — indeed, there is very little point in determining the authors of pseudonymous works if this information is not fully

[1] Note to self — find an example from the 1790s

inforporated into the scholarly record. But the particuar implementation also occludes how eighteenth century readers actually encountered author information. An identified name *overwrites* the information recorded on the title page. "Real" names are provided even in cases where this substitution contains little real information. One work, titled "Observations on the origin and effects of the Test Act" and printed in 1790, has its author listed in the ESTC as "Hudson, active 1790." A note suggests that this attribution is somewhat shaky: "Attribution from Halkett & Laing which seems to be based on a MS attribution on O copy titlepage, no longer legible. O copy with MS date May 18. 1790. Verify number and position of advertisement leaves." This is the only item in the ESTC attributed to "Hudson, 1790" so identifying "Hudson" in this way does not provide value by linking the work to another. In contrast, the title page attribution identifies the author as "a dissenter." This, I contend, is a more useful identification of the author.