

Book genre and author gender: romance>paranormal-romance to autobiography>memoir¹

Mike Thelwall, Statistical Cybermetrics Research Group, University of Wolverhampton, UK.

Although gender differences are known to exist in the publishing industry and in reader preferences, there is little public systematic evidence about them. This article uses evidence from the book-based social website Goodreads to provide a large scale analysis of 50 major English book genres based on author genders. The results show gender differences in authorship in almost all categories and gender differences the level of interest in, and ratings of, books in a minority of categories. Perhaps surprisingly in this context, there is not a clear gender-based relationship between the success of an author and their prevalence within a genre. The unexpected almost universal authorship gender differences should give new impetus to investigations of the importance of gender in fiction and the success of minority genders in some genres should encourage publishers and librarians to take their work seriously, except perhaps for most male-authored chick-lit.

Introduction

Although decreasing gender inequalities seem to be the norm in richer nations, there are some areas of life in which large differences persist. This includes employment (e.g., a high proportion of female nurses), education (few female engineers, many female behavioural scientists), social life (e.g., friendship homophily by gender) and culture (e.g., women's and men's magazines, male and female-oriented television programmes and films, e.g., Jones, 2015; Modleski, 2008). Within the realm of culture, differences in consumption are fluid and partly due to gender socialisation and commercial incentives to produce and market works that address the (presumably socialised) needs of one gender (Barletta, 2006; Fagot, Rodgers, & Leinbach, 2000; Fine, 2010; Gentry & Harrison, 2010; Vigorito & Curry, 1998).

This article focuses on one aspect of culture, book reading, and takes advantage of an online source of information about books to investigate the relationship between gender and genre in book authorship at a macro level. This is a macroscale (Jockers, 2013; Moretti, 2005) approach to the analysis of literature, using a big data approach rather than analyses of individual books. Although slightly more women than men seem to read books, and fiction in particular (see below), there seem to be large differences for some genres. In particular, there is a mainly female readership for romantic novels and there has historically been a male readership for adventure stories and science fiction (for a brief review see: Thelwall & Kousha, in press). These differences seem to be changing over time as genres and society evolve (Qi, 2007), as major events create social changes (Ugochukwu, 2011) and as influential books are published (Isaacson, 2013) but do not seem to be well understood. For example, romantic fiction typically involves a man and a woman and although the reasons why women often enjoy these stories have been extensively analysed, there does not seem to be a consensus about why men seem to be relatively uninterested. One argument is that historical association of romance with femininity may repel many men (Berlatsky, 2014). Moreover, changes in the gender readership of specific (evolving) genres,

¹ Thelwall, M. (in press). Book genre and author gender: romance>paranormal-romance to autobiography>memoir. *Journal of the Association for Information Science and Technology*.

such as science fiction, also seem to be imperfectly understood. For example, the Hunger Games trilogy is a highly successful genre-crossing science fiction adventure story with a strong romance element that has successfully attracted a young female audience (Miller, 2010).

The lack of evidence about the changing relationships between genre and gender may be due to the lack of systematic data about reader genders for book genres. Previous evidence has been mainly compiled by non-academic organisations, such as publishers and interest groups, or informally published by academics. Because of this, the raw data needed for a systematic analysis has not often found its way into mainstream academic research. Moreover, the concept of gender is itself rapidly evolving (Kacen, 2000) and so previously understood relationships between gender and genre should not be taken for granted and need ongoing sources of evidence to monitor.

Social websites like Goodreads provide a new public source of data about book genres and audiences that can be exploited to investigate the role of gender. They have a mass audience that publically records not only the books that they have read, or intend to read, but also publically rate these books and shelve them by genre. Such sites host for the first time large scale public information about the readership and reception of books. The previous best source of such information was bestseller lists and ratings of books by professional reviewers in newspapers and magazines. Although many booksellers now have online stores that allow public reviews and ratings, these do not seem to be as active as Goodreads and are potentially tainted by the direct association with sales. The biggest non-selling book information website seems to be Goodreads (owned by Amazon), which was the 121st most popular website in the USA and the 278th most popular in the world in February 2016 according to Alexa.com (<http://www.alexa.com/siteinfo/goodreads.com>). It is a social website with a majority (76%) of female users (Thelwall & Kousha, in press). Its members apparently choose the extent to which they engage in social and/or book-based activities within the site (Thelwall & Kousha, in press). Book-based activities include listing books, rating books and writing reviews. Social activities allowed within the site include registering others as friends and participating in discussion groups. Goodreads seems to be recognised as an important site for books by authors and the press (Gray, 2015; Little, 2014) and is probably used by publishers as an informal source of feedback on their offerings – and the same may be true for some academics (Zuccala, Verleysen, Cornacchia, & Engels, 2015).

Although many publisher-related sites have hosted statistics about popular books, and some surveys have analysed the preferences of readers, there do not seem to have been systematic analyses of the popularity of books on a large scale by author gender and genre, other than small scale analyses of bestsellers or literary awards (e.g., Verboord, 2012). The current paper fills this gap using data from Goodreads. This appears to be the first such use of Goodreads even though one study has analysed the site's users (Thelwall & Kousha, in press), finding that about two thirds are female and that they use a highly varied combination of social and book-based activities in the site. Another investigation used Goodreads to investigate changing reactions to books after they have won a literary prize (Kovács & Sharkey, 2014; for background, see: English, 2005) and a third showed that more books get rated on Goodreads than reviewed in the press and the increased attention particularly affects female authors (Verboord, 2011). In addition, some articles have used Goodreads reviews to help investigate individual books (Milota, 2014)

Background

In richer nations, most people probably read books, although there are gender differences. In 2014 in the USA, 77% of adult women and 67% of adult men had read a book, with higher proportions amongst younger and more educated people (Rainie & Perrin, 2015; see also: Zickuhr & Rainie, 2014). The women surveyed had read an average of 14 books in the previous year compared to 9 for men. A previous study had found that only half of American adults had read fiction in 2008, including 58% of women and 42% of men (NEA, 2009). Hence, gender differences seem to be even larger in the consumption of fiction. Women are also more frequent book readers in Germany (Süßmann, 2015) and the UK (Flood, 2013). In Europe, men spend more of their leisure time reading in Belgium, Estonia, Hungary and the UK but women spend more free time reading in Finland, Germany, Sweden and Norway, with no difference in France (Winquist, 2004). This includes reading newspapers and magazines, however, and since men may consume more news than do women (e.g., on mobile devices in the USA: Pew, 2012; and in the UK: NRS, 2015a) and women read more magazines than do men (in the UK: NRS, 2015b) it is not clear who spends the most time on reading books and fiction overall.

Gender and authorship

Books have been written by women for thousands of years (Chen, 1996), but have been overwhelmingly male-authored until a few centuries ago. The emergence of women as a numerically important body of authors appears to have occurred in the eighteenth century in the UK and USA (Prescott, 2003; Turner, 1992), with women perhaps overtaking men in the UK at about the middle of the eighteenth century after a period of rapid growth (Stanton, 1988). This growth seems to have been caused by a combination of increased female literacy (Castle, 1997) and other socio-economic changes in society that allowed women to write, the increasing availability of female role models, and an expanding cultural market (Brewer, 2004), including many women (Stanton, 1988). Despite this change, women were regarded as lacking in the judgement required to analyse literary work by authoritative male contemporaries (Castle, 1997). In the early nineteenth century in the UK most novelists were female, and by the end women wrote the most popular stories. Perhaps because of this, the literary novel emerged in the late nineteenth century as a separate, less popular but more prestigious male dominated genre (Tuchman & Fortin, 2012; for a case study, see: Corse & Westervelt, 2002; Yaeger, 1987). This continued an evolving trend of gender differences in the typical genres of books written, and women still seem to be under-recognised for the literary value of their works (Verboord, 2012). This extends to many other art fields and is an international phenomenon (Berkers, Verboord, & Weij, 2014). Such issues may be increasingly problematized in an era of ethical concerns about reviewing (Xu & Xi, 2013).

There seem to be few systematic academic statistical analyses of gender in authorship, especially for fiction, but many informally-published sources of relevant data. These informal sources typically originate from businesses, interest groups or pressure groups that monitor gender routinely or in order to support arguments about gender imbalance.

An exception to the above claim is that there are a number of macroanalysis studies of gender in authorship for specific collections of texts. For example, an investigation into a manually-curated list of 758 books by Irish-American authors over a 250 year period found

geographic differences in the proportion of female authors and substantial fluctuation in this proportion over time (Jockers, 2013, Chapter 5). An analysis of Black Drama also found substantial differences in the proportion of female authors (Jockers, 2013, Chapter 5) and so it seems that the influence of female authors changes over time and this change may be related to socio-historical factors that are specific to different communities.

For books at the top of The New York Times Adult Fiction Bestsellers List 1960-2015, 37% were written by solo female authors in comparison to 56% by solo males. The gap narrowed in about 1990 and since 2010, although the trend has continued, female-authored books have spent longer at the top of the charts (150 weeks) than have male-authored works (124 weeks) (Ortiz, 2015). In the UK in 2013, female authors sold the most bestsellers overall, especially in the children, young adult and romance categories, although men dominated the crime, thriller and adventure categories (Cherny, 2012). More recently 13 out of the 2015 top 20 selling Amazon.co.uk books were female-authored (<http://www.amazon.co.uk/gp/bestsellers/2015/books> - 5 were male authored) and from Amazon.com 6 were female-authored and 11 were male-authored (<http://www.amazon.com/gp/bestsellers/2015/books>).

Newspapers with a reputation for literary reviewing may cover male authors more than females, continuing the historical trend in which fiction written by women has been less highly regarded by authoritative literary figures. For example, 42% of New York Times book reviews from 2014 were for female authors, consistent with previous years (VIDA, 2015; see also: Weiner, 2012). In 2014, only 25% of books reviewed by the Times Literary Supplement were by women, and only 31% of reviewers were female (VIDA, 2015). Since more reviewed books seem to be more recognised (at least by libraries: Shaw, 1991), this may undermine the sales of female authors.

Men and women seem to have somewhat different writing styles because readers (Janssen & Murachver, 2005) and computers (Koppel, Argamon, & Shimoni, 2002) are able to detect the gender of the author of a work from the text itself with an above-chance level of accuracy. A core difference is that women tend to have a more involved (conversation-like) writing style and that men have a more informational (abstract academic) writing style (Biber, 1991, 1995) although this may be mainly due to the choice of subject matter for the documents analysed. Women also seem to deal with some themes in their writing more than do men, such as “affection and happiness” and “female fashion” in one analysis of 3,346 nineteenth century books (Jockers, 2013, Chapter 8).

There is a substantial body of academic research that investigates the role of gender within fiction in the sense of how it is portrayed within stories (Page, 2007). The focus here is on the authors and audiences rather than the content of the stories, however. Nevertheless, it is useful to explore two gendered fiction genres in order to gain insights into types of stories that have unbalanced author genders.

Romantic fiction

A book's genre confers expectations about its subject matter and narrative structure (Pyrhönen, 2007). Genres can be broad, such as romance and detective stories, or narrower, such as category romance and police procedural. Genres are useful for readers as a device to help them identify books that they may like, and for publishers as a way to help them identify the likely size and nature of a book's audience. For authors, conforming to a specific genre implies a degree of imitation of previous works (Pyrhönen, 2007). Nevertheless, genres are fluid and continually evolving, with boundaries that are not rigid (Todorov, 1990).

Romantic fiction is a female-oriented genre that typically involves the story of an initially difficult relationship between a woman and a man, told from the woman's perspective, and with a happy ending that often occurs after early misunderstandings are clarified through a heart-to-heart talk by the protagonists. Category romances are a subgenre of relatively short stories that are systematically published with a fixed number in each line released every month and sometimes having clear themes or styles (e.g., Harlequin Silhouette Desire; Mills & Boon Medical Romance; see also: <https://www.goodreads.com/genres/category-romance>). These novels seem to be written almost exclusively by women for women. In 2008, only 1 out of 200 Mills & Boon staff writers were male, and he used a female pen name (Jackson, 2008). An analysis by the Romantic Novelists Association with British data found that 93% of books bought with romantic content targeted female readers and such books accounted for 22% of adult fiction sales (Jones, 2009). Although this source may not be fully objective and there have been high profile historical examples of successful male-oriented romances (Bellanta, 2014), these findings seem reasonable. Category romance is often criticised for being highly formulaic (Tapper, 2014), but this unfair because, for example, poetic works are rarely criticised for adherence to formulaic elements, such as iambic pentameter or haiku. It is also an oversimplification because of the presence of sub-genres within romantic fiction and category romance. In addition, the form has evolved over time to maintain its market dominance in fiction during the current period of decline in the publishing industry (Barra, 2014; Hubbard, 1985; Tapper, 2014).

A highly successful author of romantic fiction has argued that "Romantic fiction is the only purely feminine art form. [] Romantic fiction is written by women for women, and only concerned with a view of the world through women's eyes" (Lamb, 1998). Similarly, "A romantic novel is an adult fairy story, repeating the recurring symbols and images which can explain life to a woman and satisfy a powerful need within her" (Lamb, 1998). Romantic novels seem to be consumed as a form of escapism that focus on, and perhaps allows the reader to vicariously and safely experience, the emotional journey of falling in love (Barra, 2014; for a psychological view, see also: Zunshine, 2006). An influential analysis of the reasons why many women enjoy romantic fiction argued that it provided an escapist fantasy pleasure away from (not necessarily unhappy) lives that typically involved unreciprocated responsibilities for the care and nurturing of others – often with the fictional heroes providing attention and nurturing for the heroines (Radway, 1984). Romances can also implicitly address female-specific concerns about consent in sexual relationships in complex ways despite – and even because of – their apparently simple overall narrative structure (Modleski, 2008).

Science fiction

Science fiction is a male-oriented genre, both in terms of authors and readers. It is subgenre of speculative fiction in which the focus is on the potential of science to transform future life. One informally published study analysed the gender of contributors to science fiction magazines (Connolly, 2014), which is a traditionally male genre. The proportion of female authors varied substantially by magazine, with some publishing more female-authored stories than male-authored stories. This suggests that there are either sub-genre gender preferences or editorial gender biases within the magazines. A similar study gave consistent results but found that female-authored stories were much less likely to be reviewed in magazines (Pahle, 2013). An investigation of a single speculative fiction publisher found that

males submitted more works for publication, although not in the Urban Fantasy/Paranormal Romance and Young Adult categories (Crisp, 2013). The author also claimed that “every genre publisher in the UK has female commissioning editors and 90% of the genre imprints here are actually run by women” (Crisp, 2013), suggesting that gender differences in authorship are not the fault of the publishers.

The male domination of science fiction echoes the male dominance of the (physical) sciences and engineering (see below) and might reflect a male-oriented concern with science and technology. Partly mirroring Radway (1984), science fiction could provide an escapist fantasy to men in with jobs involving technology with possibilities that are limited by practical considerations. Moreover, early influential science fiction authors were male, reflecting an overwhelming male dominance of science at the time, and their stories may have included poor roles for women and little insight into likely change in social relationships over time. For example, one story by a successful author that was published in 1955 and set in the far distant future after the earth had conquered the galaxy (Chandler, 1955) included the first female space pilot (this honour went to Valentina Tereshkova in 1963). In other stories, men in the far distant future battled aliens on strange planets and returned home to a meal prepared by a housebound wife. More recent science fiction seems to have broken male dominance somewhat, perhaps with the inclusion of elements from female dominated speculative fiction sub-genres, such as paranormal romance, changes within popular televised science fiction, such as Star Trek (Rey, 2001) and changes introduced by influential individual female authors like Ursula Le Guin (Rashley, 2007; Walsh & Le Guin, 1995).

Other types of authorship

There are gender imbalances in many professions that involve writing. For example, there are imbalances in journalism, academia and scriptwriting. Women accounted for only 36% of US newspaper newsroom staff in 2015, unchanged since 1999 (ASNE, 2015). Female authors accounted for 22% of headlines in the New York Times in January 2006 (Sobieralski, 2006).

Within academia, female authors account for 30% of journal articles globally and are in a minority in nearly all nations (Sugimoto, Lariviere, Ni, Gingras, & Cronin, 2013). This difference must be shrinking because in the USA, the proportion of doctorates awarded to females has increased rapidly from 11% in 1970 to 54% in 2010 (Di Prete, & Buchmann, 2013) so women in the USA already write the majority of doctoral theses and may even overtake males in overall academic authorship over time, although there are also differences in output volume (Sugimoto, Lariviere, Ni, Gingras, & Cronin, 2013). There do not seem to be systematic statistics about the fields in which females are more likely to be authors but the patterns presumably lag those of doctoral students. In the USA in 2014, women were formed the majority enrolled on doctoral degrees in the Arts and Humanities (51.3%), Biological and Agricultural Sciences (51.3%), Education (67.8%), and Social and Behavioural Sciences (58.3%) but were a minority in Business (44.8%), Engineering (25.4%), Mathematics and Computer Science (27.0%), and Physical and Earth Sciences (35.4%) (Allum & Okahana, 2015). The overall trend seems to be for females to be the majority in life-focused subjects, from life sciences to the arts and humanities, whereas males dominate in the rest. A simplistic characterisation of this might be that women have more interest in human relationships and nurturing whereas males tend to be more interested in making things. The reasons for imbalances are complex, however (e.g., Ceci & Williams, 2011).

There is evidence of substantial gender differences in scriptwriters. Within the US television industry, a contemporary gender imbalance in writers has been observed, with only 25% of prime time TV writers in 2014-15 being female (Lauzen, 2015a). Only 13% of the writers of the top 700 US films from 2014 had a female writer (Lauzen, 2015b), a slight increase on previous years (Smith, Choueiti, & Gall, 2010). Similarly, out of all UK independent films 2011-13, only 16% had a female writer (Steele, 2013), although the proportion was higher for more successful films in both countries.

Research questions

Following the above discussion, gender differences in authorship have been previously observed in specific subsets of romance and science fiction, but the exact prevalence across genres is unknown. It is also not known whether author gender influences the amount of interest in a book within specific genres, or whether the book reading public prefers books by any gender within specific genres. These issues form the research questions for this study.

- In which genres are there substantial gender differences in authorship?
- In which genres does the author gender influence the amount of interest shown in a book?
- In which genres does the author gender influence the popularity a book?

Methods

Data

A simple probability sample of Goodreads book pages was taken. This was then filtered to remove books that could not be used for various reasons in order to produce the final sample for analysis. Details of these steps are given below.

The Goodreads website has a sitemap that lists the URLs of all public pages. This sitemap was downloaded in January 2016 and used at the complete population of Goodreads public pages, although it may omit recent additions. A random number generator was used to select a sample of size just over 500,000 to allow for deleted pages (500,166). These randomly selected Goodreads pages were then crawled slowly (1 page every ten seconds to avoid overloading the Goodreads site) with the free web crawler SocSciBot (socscibot.wlv.ac.uk), which checks the Goodreads robots.txt file for conformity (Thelwall & Stuart, 2006). The crawl took place during January and February of 2016 and generated 500,087 valid Goodreads book pages.

Book metadata was extracted from each page using a utility added to the free software Webometric Analyst (Books tab, Get Goodreads Book Stats button). The information extracted included the name and author of the book, the number of times that it had been rated and its average user rating. Users rate books by clicking on between 1 and 5 stars in the website and so the average rating is the average number of stars given by users that rated the book. Not all books had been rated and books without any ratings were removed. It seems likely that most of these books would not have attracted an audience and so removing them is a reasonable step to focus on more recognised books. Nevertheless, many books may have received ratings only from the author or publisher. There is no way to identify this because the ratings cannot be traced to their originator using public information in the site.

Book pages sometimes also list up to 10 genres together with the number of users that have assigned a book to the given genre. This user-supplied genre information can be used to analyse book genres from the perspective of users. Although there are recognised fiction genres, these are fluid and there is no accepted authoritative list. Official categorisation systems, such as the Dewey Decimal Classification and the Library of Congress Subject Headings are also not helpful for fiction genres. User-supplied genres therefore seem to be the best solution.

Books were assigned to all genres that had at least one user in their list of up to ten genres. Many books did not have any genre assigned and these were removed, leaving a total of 201,560 books with at least one genre. Only 203 of these books did not have at least one rating, leaving 201,357 books with rating and at least one genre. This produced a subset of the initial data set, with up to 10 replications. It would also have been possible to just include the genre for each book with the most users as its main genre, but the more inclusive list is more useful because it includes additional broader genres that give context to the results. The 50 genres with the most books were selected for further analysis as the (broadly) most popular in terms of publishing, if not in terms of reading (Table 1). The top 50 was chosen in order to give a wide range of popular genres to investigate. Less popular genres may also be interesting but results about them would be less statistically powerful and so it seems reasonable to focus on the largest genres. The genres include both main genres (e.g., fiction) and sub-genres (e.g., fantasy>paranormal) as specified in the book page. For instance, a book with the *fantasy>paranormal* genre would not also be classified as *fantasy* unless this was a separate entry in the genre list.

Table 1. The 50 genres with the most books assigned, as reported in lists of (up to) ten genres in each book page. The Users column is the total of the numbers of users assigning each book to the genre.

Genre	Books	Users
fiction	41475	1218673
non-fiction	40208	507491
romance	29205	676026
fantasy	19909	1057426
history	16315	199503
children's	14147	163267
contemporary	13853	204599
mystery	13093	389375
young-adult	11286	621919
science-fiction	9967	261253
historical-fiction	9243	309406
fantasy>paranormal	9094	261909
historical	8654	137803
sequential-art>comics	8567	166331
biography	7925	103156
short-stories	7834	96615
children's>picture-books	7410	131850
romance>contemporary-romance	7403	91478
adult	7043	72240

poetry	7011	111621
adult-fiction>erotica	6981	78255
sequential-art>graphic-novels	6961	169828
sequential-art>manga	6623	285353
humor	6409	87725
reference	6039	27524
romance>m-m-romance	5729	125520
horror	5545	161636
classics	5187	664000
philosophy	5131	95606
religion	5056	54552
thriller	5003	86473
adventure	4822	83352
mystery>crime	4786	72899
novels	4564	52933
art	4469	30879
science	4463	71467
christian	4356	45478
romance>paranormal-romance	4239	110105
women's-fiction>chick-lit	4072	91559
politics	3894	34030
suspense	3829	41560
romance>historical-romance	3767	103730
literature	3697	77384
autobiography>memoir	3673	67055
food-and-drink>cookbooks	3642	36381
animals	3280	29674
psychology	3259	49520
thriller>mystery-thriller	3167	26621
fantasy>magic	3028	60821
travel	2941	31811

For each genre, a list was built of all books assigned to it by users. Goodreads does not report the gender or nationality of authors in book pages but in most cases author (apparent) gender can be inferred from their first name. This process is culture-specific and since most authors seemed to be American, or at least from an Anglophone nation, the books were restricted to those written in English to reduce, but not eliminate, the number of authors with first names with a different Anglophone gender (e.g., Andrea, Kim, Nicola). US census data from 1990 was then used to select first names that were (a) in the 10000 most popular names in the US, (b) had at least 4 letters (to avoid confusion with initials or short words –e.g., A.S Byatt), and (c) was used at least 90% of the time by one gender. US census data seems to be optimal for this because the USA incorporates people from many different national origins and so its data is probably more culturally robust for gender than the data from most other countries. The result was a list of 4772 gendered first names, from Aaron (M) to Zulma (F). This list was then used to assign a gender to authors via their first

name with an overall probability of over 90%. Books from authors with first names that were not in the list were removed. Authors with gender swapping names will be misclassified (e.g., George Eliot, Robert Galbraith) as will authors that use gender swapping versions of their first name (e.g., Ali Smith). The gender analysis is restricted to these two genders because of the lack of effective automatic techniques for finding other genders (e.g., transgender authors Eddie Izzard and Caroline Cossey) and because their relative scarcity blocks statistically significant findings.

Analyses

The prevalence of books by author gender for each genre was obtained by simply counting up the number of matching books in each category.

The level of interest in a book for both genders and all genres was estimated by calculating the geometric mean number of ratings per book. A rating for a book implies that a person has read at least part of a book and it seems to be the best available indicator of interest in a book from the Goodreads site. The main alternative, the number of reviews of a book, also suggests reading a book but probably reflects a deeper interest as well as a desire to write about it. It would therefore be a less clear interest indicator because the wish to write a review may be associated with other goals, such as language learning, literary educational tasks, or the desire to communicate with others. It would also be possible to total up the number of lists that a book is on (e.g., read, to read) but this also seems less straightforward because a user may place a book that they have read on multiple lists or may not list it. The geometric mean was used rather than the arithmetic mean because the rating counts data sets were highly skewed. The geometric mean was calculated as the arithmetic mean of the sum of the ratings+1, with the result transformed back with the function $\exp(x)-1$, following standard practice. Confidence intervals were calculated using the t-distribution for the log-transformed data, which was approximately normal.

The average ratings for a book for both genders and all genres was calculated as above except that standard arithmetic mean was used instead of the geometric mean because the data was not highly skewed.

The results are presented separately for the top 25 genres and the next 25 genres in order to avoid small font sizes in the graph labels.

Results

There are substantial gender differences in authorship in some categories, such as romance where nearly all authors are (apparently) female and sequential-art>comics, where male authors dominate (figures 1, 2). More generally, there seems to be reasonably substantial gender difference in authorship in all categories except *autobiography>memoir*. Female authors dominate all the romance categories, including even male-male romances (romance>m-m-romance). This is an English-language parallel (Berlatsky, 2014) to the Japanese Yaoi phenomenon of male gay manga written by women for women (Mizoguchi, 2003; Suzuki, 1998). Male authors are in the majority in most other categories, except for children and adult (presumably meaning erotica) categories, fantasy, suspense and cook books.

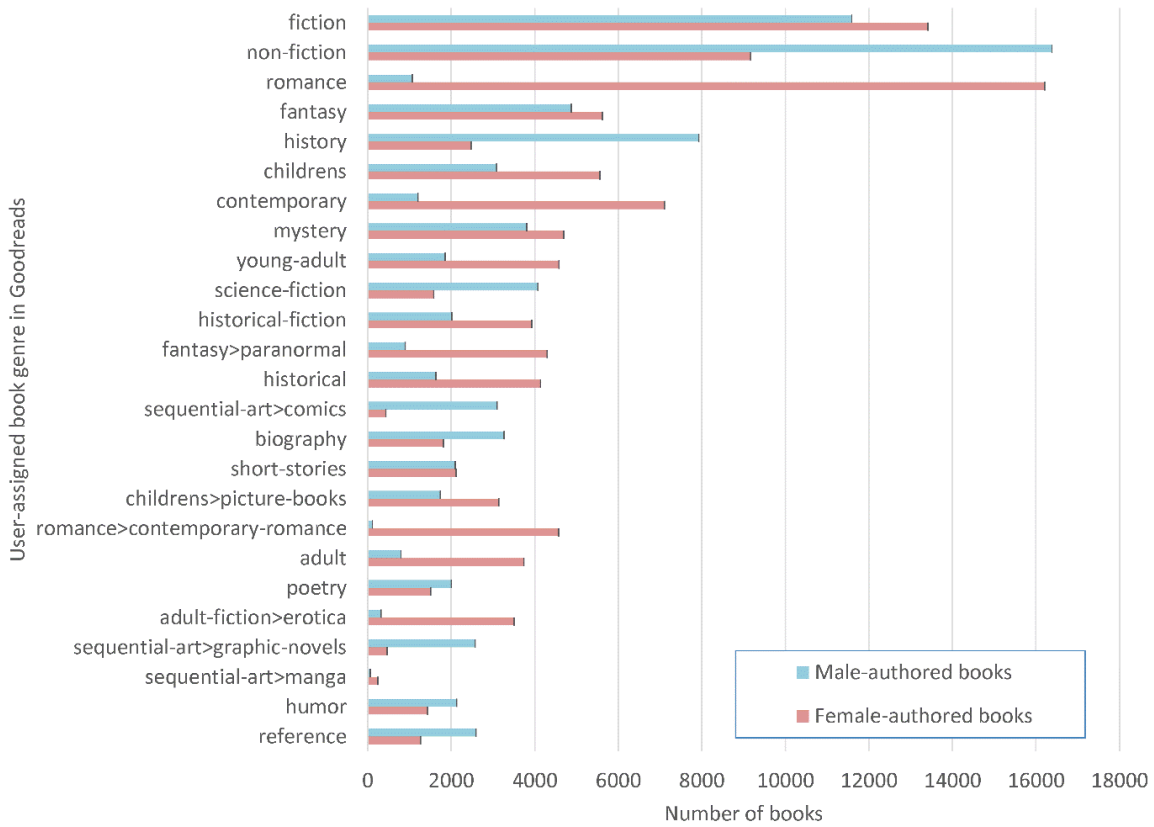


Figure 1. The number of books by male and by female authors in the top 25 genres from the crawled sample.

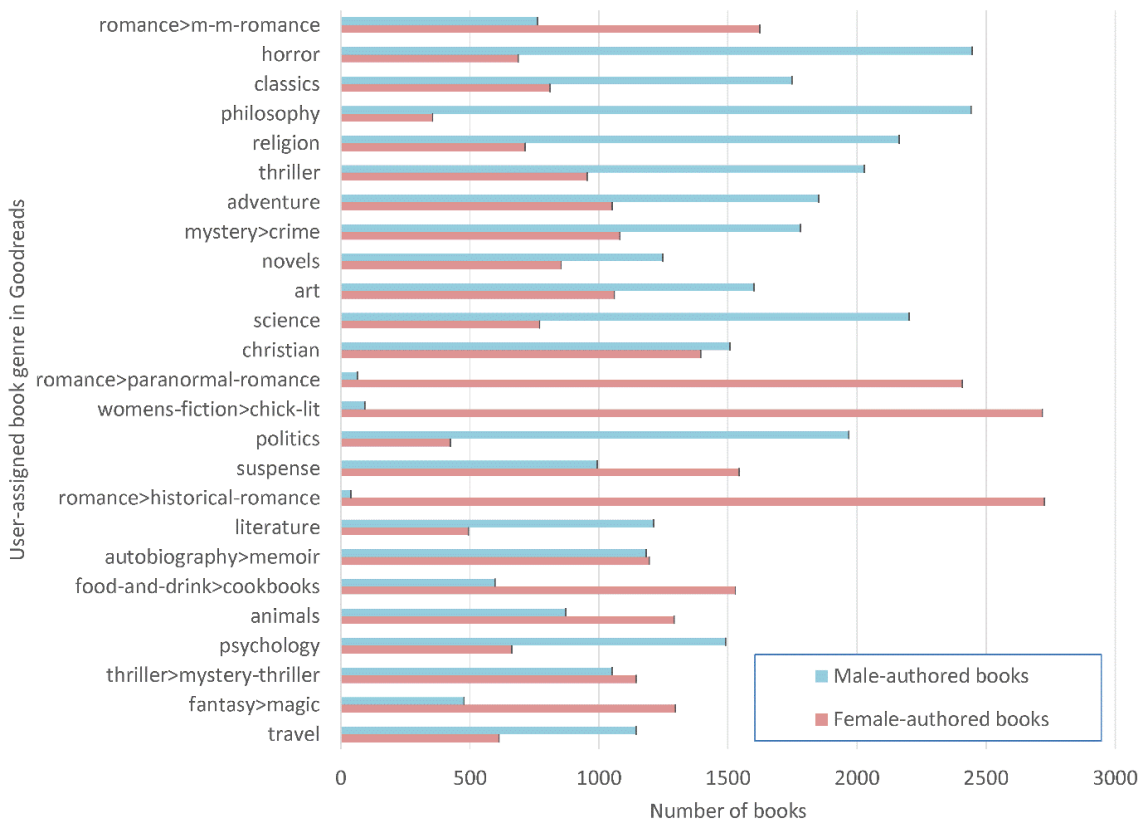


Figure 2. The number of books by male and by female authors in genres 26-50 from the crawled sample.

The results for the average number of reviews per book reveal small gender differences in most cases (figures 3, 4), in comparison to the large differences in authorship discussed above. Most of the differences are within the range of sampling variation expressed by the confidence intervals, with a few exceptions.

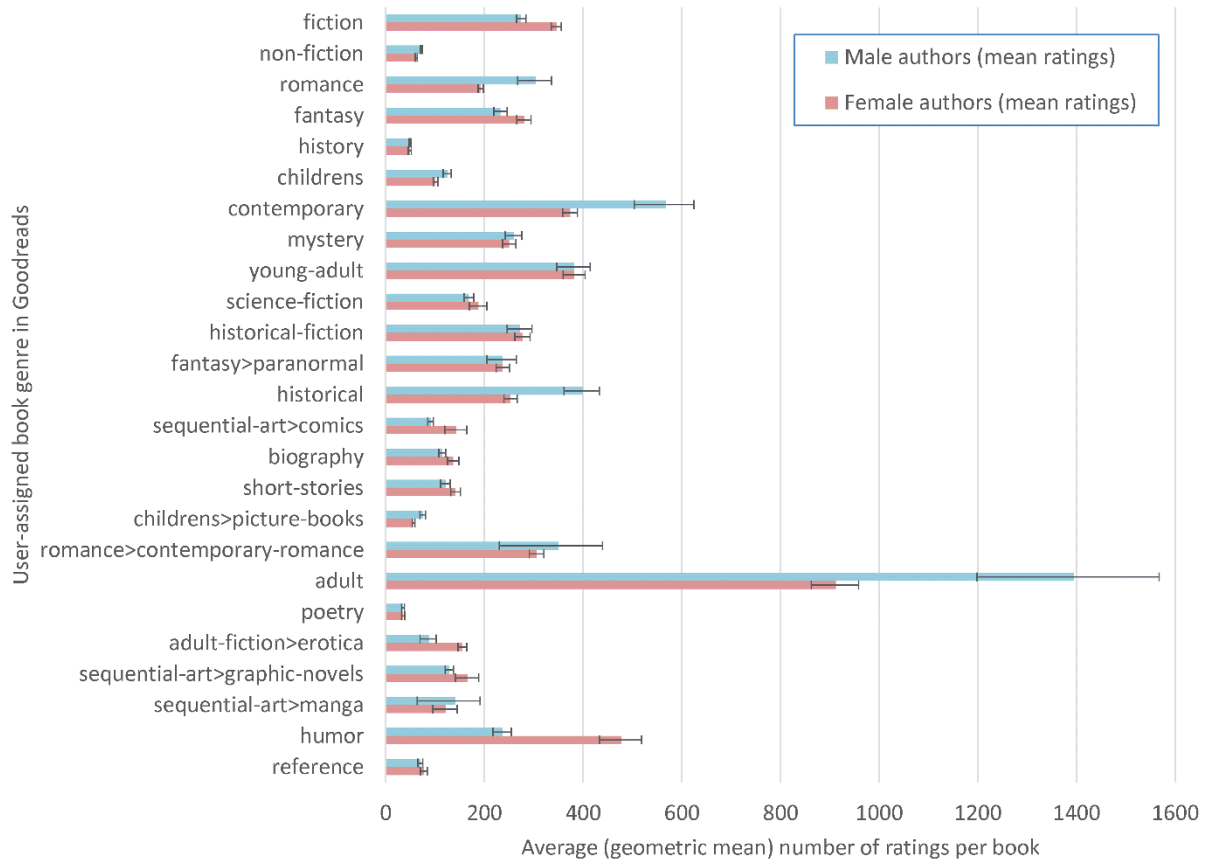


Figure 3. The number of ratings per book for male and by female authors in the top 25 genres from the crawled sample.

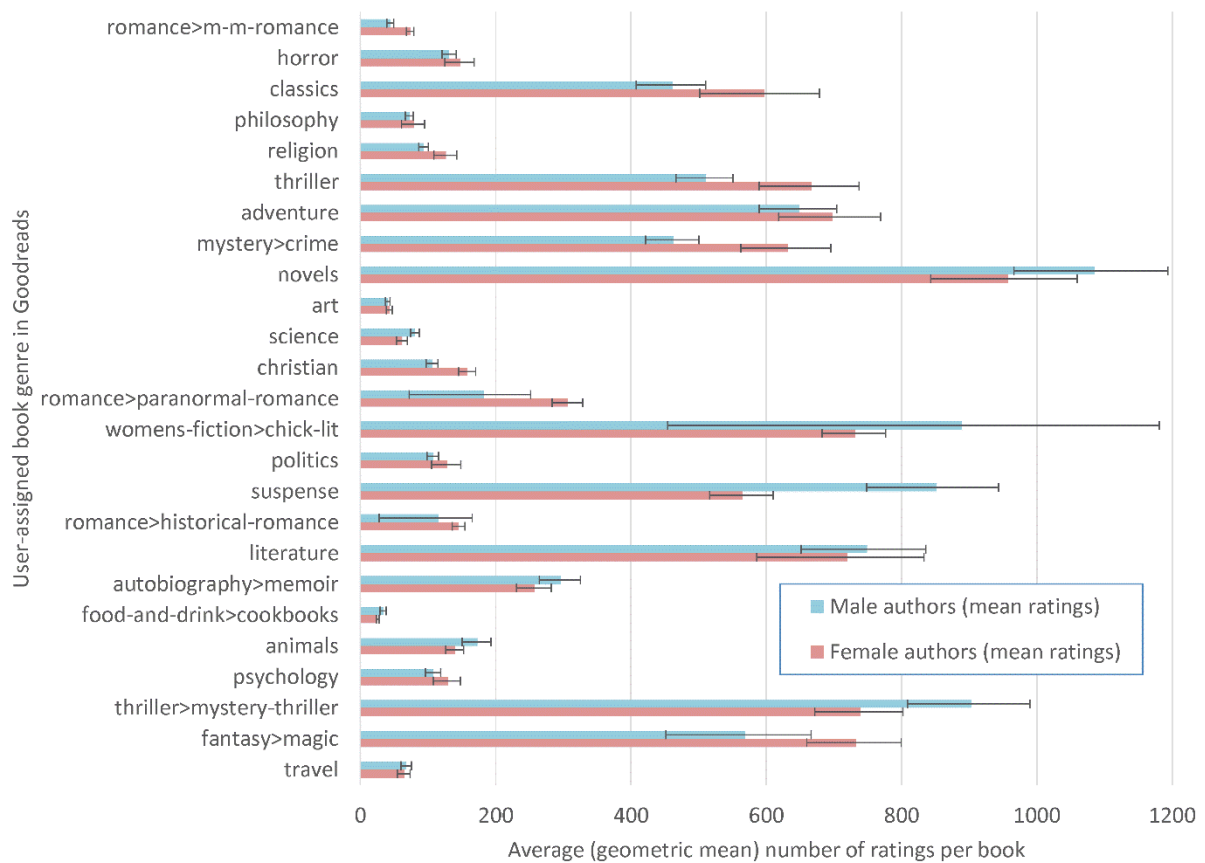


Figure 4. The number of ratings per book for male and by female authors in genres 26-50 from the crawled sample.

The ratings given to books tend to be quite positive, with all genres having an average of between 3.5 and 4.1 stars (this is 15% of the maximum theoretical 1-5 range). Poetry, art and Christian genres have the highest average. The gender differences within genres are again relatively small compared to the author gender differences.

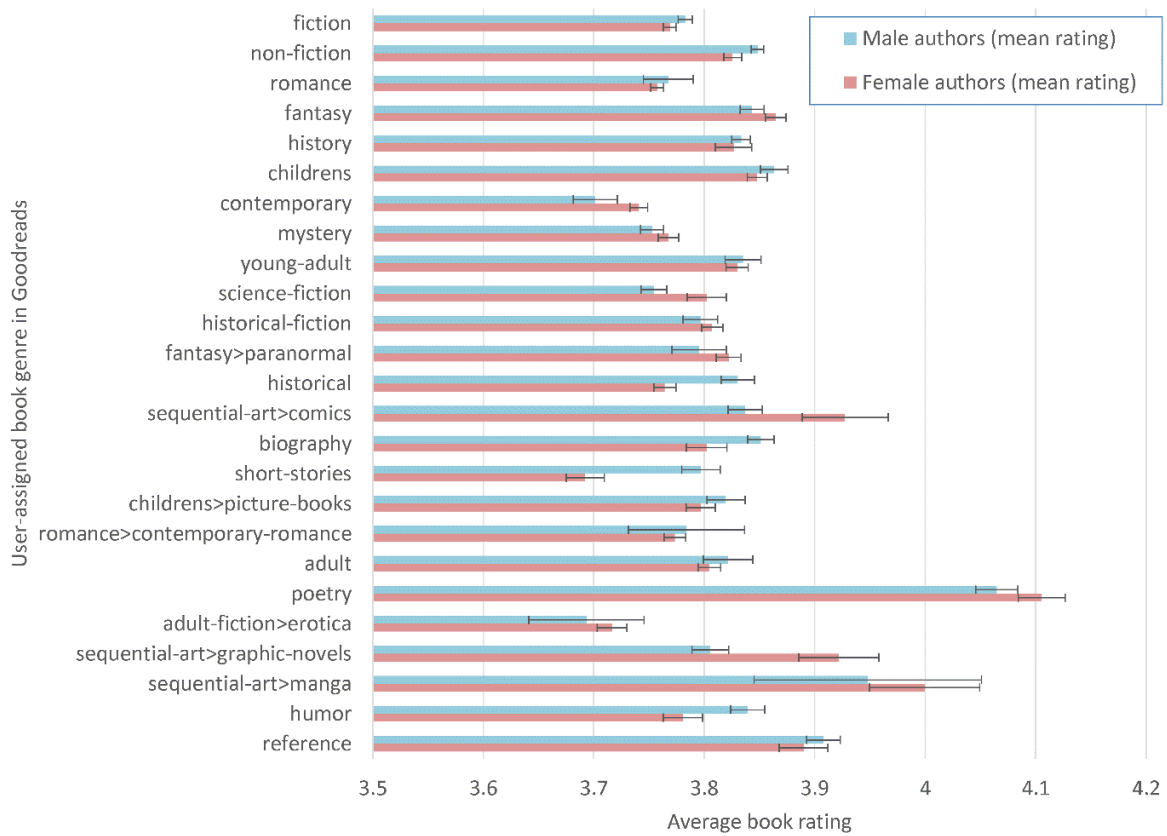


Figure 5. The average rating for books by male and by female authors in the top 25 genres from the crawled sample.

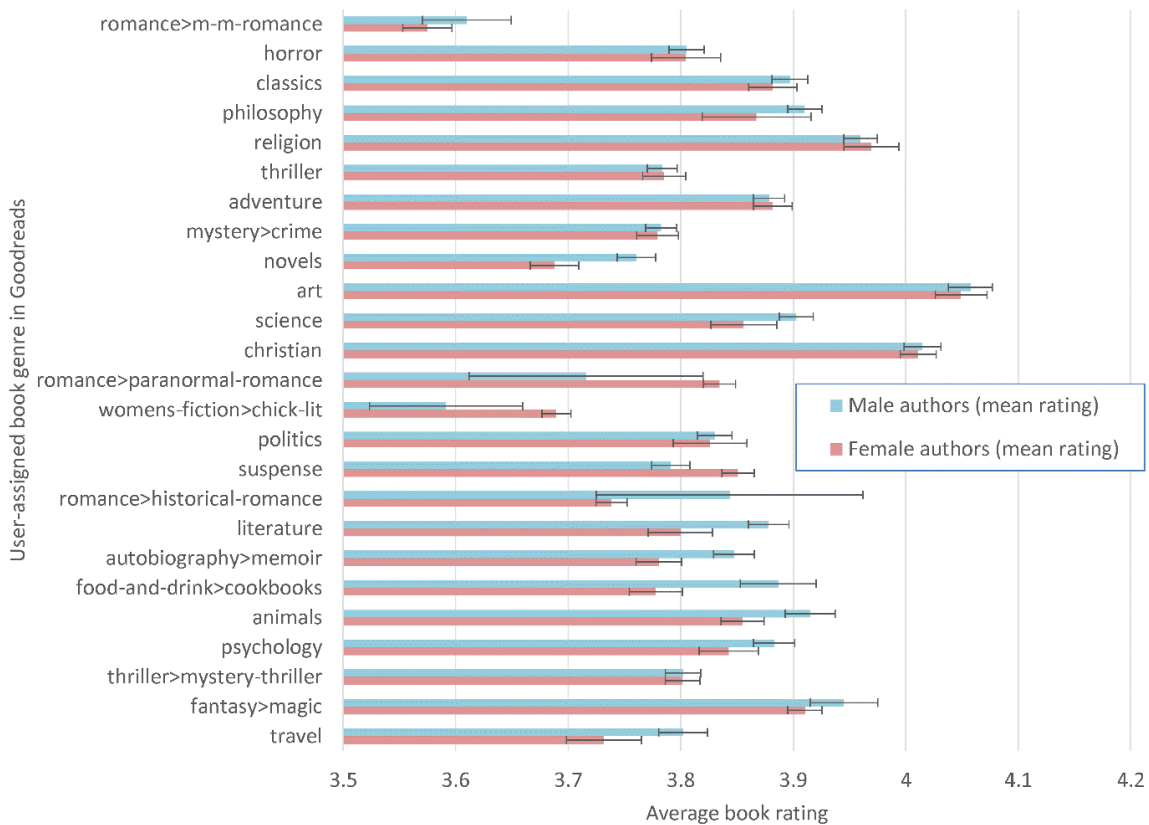


Figure 6. The average rating for books by male and by female authors in genres 26-50 from the crawled sample.

Discussion

A limitation of the analysis is that Goodreads' 76% female user base is likely to give a gender-biased reception to the books analysed in terms of their popularity and ratings. In addition, there is a heavy US-bias, and a restriction to internet users. People that have joined Goodreads are probably also a special case of internet users with a particularly strong interest in books. They are presumably more active book readers and may contain a disproportionate number of people with additional interests in books, such as through studying literature or working in the publishing industry. The gender identification method is also a limitation because some authors use pseudonyms, perhaps especially to hide their gender when they are in a minority in a category. It is also likely to under-represent the work of ethnic minorities in English because these might not be frequent enough to appear in the first name lexicon used, and may omit ethnic minorities that have non-gendered first names (e.g., Sikh). Another limitation is that prolific authors with multiple books in Goodreads could be particularly influential in the findings. Presumably the reception and readership of the books of a single author would be broadly similar and hence bias the statistical findings here (because the data for multiple books by the same author is unlikely to be statistically independent). Although it would have been possible to analyse a maximum of one book per author, this might give misleading findings by underrepresenting the role of prolific, and presumably successful, authors. Nevertheless, the potential influence individual prolific authors should be taken into account when interpreting the results.

In answer to the first research question, there are gender differences in authorship in almost all genres and the differences are substantial in most of them. The main exception is the short story genre, which is perhaps surprising given the recent prominence of women writers, such as Alice Munro, whose Nobel Prize citation was "the master of the contemporary short story" (Nobel, 2013). Although gender differences in the romance and science fiction authorship were previously known, and substantial gender differences in non-book writing professions are also known, the widespread existence of differences across genders is apparently a new finding.

In answer to the second research question, in a minority of genres there are gender statistically significant gender differences in the interest shown in books, as follows.

- Men attract more ratings for non-fiction, romance, children's, contemporary, historical, children's picture-books, adult, thriller, science, suspense, thriller, mystery-thriller.
- Women attract more ratings for fiction, fantasy, sequential-art, comics, adult-fiction, erotica, humor, romance, m-m-romance, religion, mystery, crime, christian, romance, paranormal-romance.

It is clear from these results that the overall level of interest in a book within a genre does not have a simple relationship with the prevalence of books by a particular author gender. For example, whilst romances are overwhelmingly female-authored, male-authored romances tend to attract more ratings. Conversely, male-male romances are mainly female-authored and male-authored books in this category tend to receive fewer ratings. A possible explanation for the romance category is that the novelty value of the author gender may in itself attract interest.

In answer to the third research question, there are gender differences in the overall ratings given to a book within specific genres, although in most cases the differences are not statistically significant.

- Men attract higher ratings for non-fiction, historical, biography, short-stories, humor, novels, science, literature, autobiography>memoir, food-and-drink>cookbooks, animals, travel.
- Women attract higher ratings for contemporary, science-fiction, sequential-art>comics, sequential-art>graphic-novels, women's-fiction>chicklit, suspense.

More positive ratings for a gender do not always associate with prevalence in the genre. For example, the few females writing science fiction, graphic novels or comics attract relatively high ratings for their work. Conversely, chicklit is overwhelmingly female authored and the few male authors get lower ratings for their work. A possible explanation for gender minority authors tending to create more highly rated work within some genres is that publishers may have a predisposition for the gender of authors for particular genres simply because they have produced more or better work in the past. If this is true (it seems unlikely from the Background discussion), then a woman might have to be better than a man at creating graphic novels in order to get them published. Alternatively, minority gender authorship may in itself attract attention in genres such as romance and comic books and this may generate either lower expectations or better ratings due to the novelty value of the author gender in some cases.

Conclusions

The results confirm that gender differences exist in authorship within genres for English language books. They extend what has been previously known by showing that such differences are the norm within genres rather than restricted to a few cases. The results show, apparently for the first time, that there are also different levels of interest in, and appreciation of, books within some genres depending on the author gender. The relationship between the three (authorship, interest, appreciation) is not straightforward, with minority genders getting higher or lower average ratings, depending on the genre.

For publishers, authors, and librarians (see also: Stevens & Sheehan, 2015; Trott & Naik, 2012), these findings might confirm and extend what they already suspect about authorship trends, but the lack of a simple relationship between gender prevalence and book popularity should give encouragement for gender minority authors, except perhaps for men writing chicklit. This is especially true given that genres are continually evolving and there are precedents for minority genders changing a genre (e.g., Ursula Le Guin's science fiction exploring gender).

For researchers that are interested in gender in literature, the findings should provide useful background context. Although they do not give insights into the *causes* behind the differences found, evidence of the *prevalence* of a phenomenon is important knowledge for those seeking to understand its nature.

For publishers, researchers, and interest groups the Goodreads website is a valuable new source of information about the reception of books that can be exploited in the future to monitor book popularity, genre changes and gender issues. Whilst publishers and authors may already see the site as a useful source of feedback about individual books and a marketing tool, this study shows that can also yield large scale information about overall trends. Perhaps future work can monitor changes in Goodreads over time and look for deeper insights into the interaction between gender and authorship using additional information found within the site. This may give an interesting complement to macroscale text-based analyses of the contents of large numbers of books of different genres and author genders (e.g., Jockers, 2013).

References

- Allum, J. & Okahana, H. (2015). Graduate enrollment and degrees: 2004 to 2014. Washington, DC: Council of Graduate Schools.
- ASNE (2015). Newsroom Census. <http://asne.org/content.asp?pl=140&sl=143&contentid=143>
- Barletta, M. (2006). Marketing to women: How to understand, reach, and increase your share of the world's largest market segment (2 ed). Chicago, Il: Dearborn Trade Publishing.
- Barra, A. C. (2014). Beyond the bodice ripper: innovation and change in the romance novel industry (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick).
- Bellanta, M. (2014). A Masculine romance: The sentimental bloke and Australian culture in the war-and early interwar years. *Journal of Popular Romance Studies*, 4(2), 1-20.
- Berkers, P., Verboord, M., & Weij, F. (2014). Genderongelijkheid in de dagbladberichtgeving over kunst en cultuur. *Sociologie*, 10(2), 124-146.
- Berlatsky, B. (2014). Why don't men read romance novels? *Pacific Standard*. <http://www.psmag.com/books-and-culture/dont-men-read-romance-novels-misogyny-femininity-publishing-books-92921>
- Biber, D. (1991). Variation across speech and writing. Cambridge, UK: Cambridge University Press.
- Biber, D. (1995). Dimensions of Register variation: A cross-linguistic comparison. Cambridge, UK: Cambridge University Press.
- Brewer, J. (2004). The pleasures of the imagination: English culture in the eighteenth century. London: Harper Collins.
- Castle, T. (1997). Women and Literary Criticism. In: Nisbet, H. & Rawson. C. (Eds) *The Cambridge History of literary criticism IV: The eighteenth century*. Cambridge, UK: Cambridge University Press (pp. 434-455).
- Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8), 3157-3162.
- Chandler, A. B. (1955). *The road to the rim*. New York, NY: Ace.
- Chen, Y. S. (1996). The Historical Template of Pan Chao's "Nū Chieh". *T'oung Pao*, 229-257.
- Cherny, L. (2012). UK bestsellers: Remash by genre and gender. <http://blogger.ghostweather.com/2012/08/uk-bestsellers-remash-by-genre-and.html>
- Connolly, S. (2014). The issue of gender in genre fiction: A detailed analysis. http://clarkesworldmagazine.com/connolly_06_14/
- Corse, S. & Westervelt, S.D. (2002). Gender and literary valorization: The awakening of a canonical novel. *Social Perspectives*. 45(2). 139-161.
- Crisp, J. (2013). Sexism in genre publishing: A publisher's perspective. <http://www.torbooks.co.uk/blog/2013/07/10/sexism-in-genre-publishing-a-publishers-perspective>
- Di Prete, T. A., & Buchmann, C. (2013). *The rise of women: the growing gender gap in education and what it means for American schools*. Miami, FL: University of Miami.
- English, J.F. (2005). *The Economy of prestige: Prizes, awards and the circulation of cultural value*. Cambridge, MA: Harvard University Press.
- Fagot, B. I., Rodgers, C. S., & Leinbach, M. D. (2000). Theories of gender socialization. The developmental social psychology of gender. In Eckes, T. & Trautner, H. (eds), *The*

- Developmental Social Psychology of Gender. New York, NY: Taylor & Francis (pp. 65-89).
- Fine, C. (2010). *Delusions of gender: The real science behind sex differences*. New York, NY: W. W. Norton
- Flood, A. (2013). Four million UK adults never read books for pleasure. *The Guardian*. <http://www.theguardian.com/books/2013/feb/14/4-million-uk-adults-never-read-books>
- Gentry, J., & Harrison, R. (2010). Is advertising a barrier to male movement toward gender change?. *Marketing Theory*, 10(1), 74-96.
- Gray, B.C. (2015). Why I quit Goodreads (or, The booknet is not safe for women). <http://bookriot.com/2015/03/25/booknet-not-safe-women-quit-goodreads/>
- Hubbard, R. C. (1985). Relationship styles in popular romance novels, 1950 to 1983. *Communication Quarterly*, 33(2), 113-125.
- Isaacson, N. (2013). Science Fiction for the Nation: Tales of the Moon Colony and the Birth of Modern Chinese Fiction. *Science Fiction Studies*, 40(1), 33-54.
- Jackson, P. (2008). Can a man really write a Mills & Boon? *BBC News*. <http://news.bbc.co.uk/1/hi/7710508.stm>
- Janssen, A., & Murachver, T. (2005). Readers' perceptions of author gender and literary genre. *Journal of Language and Social Psychology*, 24(2), 207-219.
- Jockers, M. L. (2013). *Macroanalysis: digital methods and literary history*. Urbana, IL: University of Illinois Press.
- Jones, B. (2015). My Little Pony, tolerance is magic: Gender policing and Brony anti-fandom. *The Journal of Popular Television*, 3(1), 119-125.
- Jones, C. (2009). Is romantic fiction popular? *Romantic Novelists Association*. http://www.romanticnovelistsassociation.org/index.php/news/entry/is_romantic_fiction_popular
- Kacen, J. J. (2000). Girrrl power and boyyy nature: the past, present, and paradisa future of consumer gender identity. *Marketing Intelligence & Planning*, 18(6/7), 345-355.
- Koppel, M., Argamon, S., & Shimon, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 17(4), 401-412.
- Kovács, B., & Sharkey, A. J. (2014). The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative Science Quarterly*, 59(1), 1-33.
- Lamb, C., (1998). Why do women read romantic fiction? <http://www.independent.co.uk/arts-entertainment/literary-notes-why-do-women-read-romantic-fiction-1176244.html>
- Lauzen, M. (2015a). Boxed in: Employment of behind-the-scenes and on-screen women in 2014-15 prime-time television. http://womenintvfilm.sdsu.edu/files/2014-15_Boxed_In_Report.pdf
- Lauzen, M. (2015b). Women and the big picture: behind-the-scenes employment on the top 700 films of 2014. http://womenintvfilm.sdsu.edu/files/2014_Women_and_the_Big_Picture_Report.pdf
- Little, J. (2014). On the importance of pseudonymous activity. <http://dearauthor.com/features/essays/on-the-importance-of-pseudonymous-activity/>
- Miller, L. (2010). Fresh hell: what's behind the boom in dystopian fiction for young readers? *The New Yorker*, June 14 & 21. <http://www.newyorker.com/magazine/2010/06/14/fresh-hell-2>

- Milota, M. (2014). From “compelling and mystical” to “makes you want to commit suicide”: Quantifying the spectrum of online reader responses. *Scientific Study of Literature*, 4(2), 178-195.
- Mizoguchi, A. (2003). Male-male romance by and for women in Japan: A history and the subgenres of “yaoi” fictions. *US-Japan Women's Journal*, 49-75.
- Modleski, T. (2008). *Loving with a vengeance: Mass produced fantasies for women* (2 ed). London, UK: Routledge.
- Moretti, F. (2005). *Graphs, maps, trees: abstract models for a literary history*. London: Verso.
- NEA (2009). *Reading on the rise: A new chapter in American literacy*. National Endowment for the Arts. <https://www.arts.gov/sites/default/files/ReadingonRise.pdf>
- Nobel (2013). The Nobel Prize in literature 2013. http://www.nobelprize.org/nobel_prizes/literature/laureates/2013/
- NRS (2015a). NRS readership estimates - newspapers and supplements. http://www.nrs.co.uk/downloads/pdf/newspapers_201509.pdf
- NRS (2015b). Magazines. <http://www.nrs.co.uk/latest-results/facts-and-figures/magazines-factsfigs/>
- Ortiz, M. H. (2015). The authors. <http://martinhillortiz.blogspot.co.uk/2015/05/new-york-times-best-sellers-look-at.html>
- Page, R. (2007). Gender. In Herman, D. (ed.) *The Cambridge companion to narrative*. Cambridge, UK: University of Cambridge Press (pp. 189-202).
- Pahle, R. (2013). Are you a female SF&F writer? You might be able to get published, but good luck getting reviewed. <http://www.themarysue.com/genre-publishing-gender-reviews/>
- Pew (2012). News consumption on mobile devices. <http://www.journalism.org/2012/12/11/news-consumption-mobile-devices/>
- Prescott, S. (2003). *Women, authorship and literary culture, 1690-1740*. Basingstoke, UK: Palgrave Macmillan.
- Pyrhönen, H. (2007). Genre. In Herman, D. (ed.) *The Cambridge companion to narrative*. Cambridge, UK: University of Cambridge Press (pp. 109-123).
- Qi, S. (2007). Traditions and trends in western literary genre studies. *Foreign Literature Studies*, 29(2), 50-61.
- Radway, J. A. (1984). *Reading the romance: Women, patriarchy, and popular literature*. Chapel Hill: Univ of North Carolina Press.
- Rainie, L. & Perrin, A. (2015). Slightly fewer Americans are reading print books, new survey finds. Pew Research Center. <http://www.pewresearch.org/fact-tank/2015/10/19/slightly-fewer-americans-are-reading-print-books-new-survey-finds/>
- Rashley, L.H. (2007). Revisioning gender: Inventing women in Ursula K. Le Guin's nonfiction. *Biography*, 30(1), 22-47.
- Rey, J. M. (2001). Changing gender roles in popular culture. In: Biber, D. & Conrad, S. (Eds.) *Variation in English: Multi-dimensional studies*. Abingdon, UK: Routledge (pp. 138-156).
- Smith, S. L., Choueiti, M., & Gall, S. (2010). *Gender inequality in popular films: examining on screen portrayals and behind-the-scenes employment patterns in motion pictures released between 2007–2009*. Los Angeles, CA: Annenberg School for Communication & Journalism.

- Shaw, D. (1991). An analysis of the relationship between book reviews and fiction holdings in OCLC. *Library and Information Science Research*, 13(2), 147-154.
- Stevens, J., & Sheehan, S. (2015). Creating a popular romance collection in an academic library. *Journal of Popular Romance Studies*, 5(1), 1-23.
- Sobieralski, C. (2006). The New York Times data mapping project. http://www.pelefire.com/nytimes/text/nytimes_data_jan06_final.htm
- Stanton, J. P. (1988). Statistical profile of women writing in English from 1660 to 1800. In: Keener, F. M., & Lorsch, S. E. (1988). *Eighteenth-century Women and the Arts*. New York: Greenwood (pp. 247-254).
- Steele, D. (2013). Succes de plume? Female screenwriters and directors of UK films, 2010–2012. BFT. <http://www.bfi.org.uk/sites/bfi.org.uk/files/downloads/bfi-report-on-female-writers-and-directors-of-uk-films-2013-11.pdf>
- Sugimoto, C. R., Lariviere, V., Ni, C. Q., Gingras, Y., & Cronin, B. (2013). Global gender disparities in science. *Nature*, 504(7479), 211-213.
- Süßmann, I. (2015). Germany: A nation of readers? <http://publishingperspectives.com/2015/07/germany-a-nation-of-readers/>
- Suzuki, K. (1998). Pornography or therapy? Japanese girls creating the Yaoi phenomenon. In: Inness, S. (ed.), *Millennium girls: Today's girls around the world*. New York, NY: Rowman & Littlefield (pp. 243-67).
- Tapper, O. (2014). Romance and innovation in twenty-first century publishing. *Publishing Research Quarterly*, 30(2), 249-259.
- Thelwall, M. & Kousha, K. (in press). Goodreads: A social network site for book readers. *Journal of the Association for Information Science and Technology*.
- Thelwall, M. & Stuart, D. (2006). Web crawling ethics revisited: Cost, privacy and denial of service. *Journal of the American Society for Information Science and Technology*, 57(13), 1771-1779.
- Todorov, T. (1990). *Genres in discourse*. Cambridge, UK: Cambridge University Press.
- Trott, B., & Naik, Y. (2012). Finding good reads on Goodreads. *Reference & User Services Quarterly*, 51(4), 319-323.
- Tuchman, G., & Fortin, N. E. (2012). *Edging women out: Victorian novelists, publishers and social change*. London, UK: Routledge.
- Turner, C. (1992). *Living by the pen: women writers in the eighteenth century*. London: Routledge.
- Ugochukwu, F. (2011). A lingering nightmare: Achebe, Ofoegbu, and Adichie on Biafra. *Matatu-Journal for African Culture and Society*, 39(1), 253-272.
- VIDA (2015). The 2014 VIDA count. <http://www.vidaweb.org/2014-vida-count/>
- Vigorito, A. J., & Curry, T. J. (1998). Marketing masculinity: Gender identity and popular magazines. *Sex Roles*, 39(1-2), 135-152.
- Verboord, M. (2011). Cultural products go online: Comparing the internet and print media on distributions of gender, genre and commercial success. *Communications*, 36(4), 441-462.
- Verboord, M. (2012). Female bestsellers: A cross-national study of gender inequality and the popular–highbrow culture divide in fiction book production, 1960–2009. *European Journal of Communication*, 27(4), 395-409.
- Walsh, W., & Le Guin, U. (1995). I am a woman writer; I am a western writer: an interview with Ursula Le Guin. *Kenyon Review*, 17(3/4) 192-205.

- Weiner, J. (2012). Why isn't this woman smiling?
<http://jenniferweiner.blogspot.co.uk/2012/01/back-in-summer-of-2010-some-female.html>
- Winqvist, K. (2004). How Europeans spend their time. Everyday life of women and men. Rapport de la Commission Européenne, Luxembourg: Pocketbooks Edition.
- Xu, Y., & Xi, Y. (2013). Ethical literary criticism: Its openness, inclusiveness and vitality. *Foreign Literature Studies*, 35(6), 171-176.
- Yaeger, P. S. (1987). "A language which nobody understood" Emancipatory strategies in *The Awakening*, *Novel*, 20(3), 197-219.
- Zickuhr, K. & Rainie, L. (2014). A snapshot of reading in America in 2013. Pew Research Center. <http://www.pewinternet.org/2014/01/16/a-snapshot-of-reading-in-america-in-2013/>
- Zuccala, A. A., Verleysen, F. T., Cornacchia, R., & Engels, T. C. (2015). Altmetrics for the humanities: Comparing Goodreads reader ratings with citations to history books. *Aslib Journal of Information Management*, 67(3), 320-336.
- Zunshine, L. (2006). *Why we read fiction: Theory of mind and the novel*. Columbus, OH: Ohio State University Press.