

Boys write about boys

Hsiao, Yaling; Banerji, Nilanjana; Nation, Kate

DOI:

[10.1111/cdev.13623](https://doi.org/10.1111/cdev.13623)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Hsiao, Y, Banerji, N & Nation, K 2021, 'Boys write about boys: androcentrism in children's reading experience and its emergence in children's own writing', *Child Development*, vol. 92, no. 6, pp. 2194-2204.
<https://doi.org/10.1111/cdev.13623>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Boys Write About Boys: Androcentrism in Children's Reading Experience and Its Emergence in Children's Own Writing

Yaling Hsiao 
University of Oxford

Nilanjana Banerji
Oxford University Press

Kate Nation
University of Oxford

Gender bias exists in our language environment. We investigated personal name usage in two large corpora of language written for and by U.K. children aged 5–13. Study 1 found an overrepresentation of male names in children's books, largely attributable to male authors. In stories written by over 100,000 children, Study 2 found an overall male bias that interacted with age. Younger children wrote more about their own gender. With age, girls became more balanced yet boys continued to show a strong male bias. Our findings demonstrate a male-centered bias in both children's books and their own writing. We consider the power of written language to both shape and be shaped by cultural stereotypes via systematic biases in gender associations.

Children are susceptible learners, absorbing information and forming concepts from a variety of sources. Text is often a vehicle for this. Well before they can read themselves, children experience vast amounts of written language via being read to (e.g., Logan, Justice, Yumus, & Chaparro-Moreno, 2019). Even books written for preschool children provide exposure to language and information well beyond the everyday "here and now" (Montag, Jones, & Smith, 2015). Once children are literate, exposure to written language grows further and school-age children rapidly acquire information from books (Hopkins & Weisberg, 2017; Young, Moss, & Cornwell, 2007). A well-documented feature of children's literature is that it contains gender stereotypes (e.g., Lewis, Borkenhagen, Converse, Lupyan, & Seidenberg, 2020). In this article, we extend these findings

by investigating whether gender biases exist in the language children themselves write.

The scale of gender biases in materials written for children is substantial. These biases are substantiated in two major ways: in content and in frequency. Conceptual associations formulate gender stereotypes (e.g., men depicted as more outgoing and intelligent, and women as domestic and obedient), whereas a high frequency of male-related terms leads to androcentrism—the overrepresentation of men and underrepresentation of women (e.g., men seen as representing humanity, men = people; Bailey & LaFrance, 2017; Bailey, LaFrance, & Dovidio, 2019; Hyde, 1984). Both of these bias types are found in children's books. Weitzman, Eifler, Hokada, and Ross (1972) analyzed the content of popular and prize-winning picture books for children and reported striking levels of female underrepresentation. Fewer female names occurred in book titles and female characters tended to play less central roles, and appeared less often in the pictures. These observations stand historically across the twentieth century (McCabe, Fairchild, Grauerholz, Pescosolido, & Tope, 2011) and although there are some

The Oxford Children's Corpus is a growing database of writing for and by children developed and maintained by Oxford University Press for the purpose of children's language research. The work for this article was supported by the British Academy Post-Doctoral Fellowship (PF2/180013) awarded to Yaling Hsiao, and grants from The Leverhulme Trust (RPG-2015-070) and the Nuffield Foundation (EDO/43392) to Kate Nation, and resources made available to Nilanjana Banerji by the Department of Children's Dictionaries and Children's Language Data at Oxford University Press. Data and code associated with this article are available on the Open Science Framework Web site (<http://osf.io/m9arj>).

Correspondence concerning this article should be addressed to Yaling Hsiao, Department of Experimental Psychology, Anna Watts Building, University of Oxford, Oxford OX2 6GG, United Kingdom. Electronic mail may be sent to yaling.hsiao@psy.ox.ac.uk.

© 2021 The Authors. *Child Development* published by Wiley Periodicals LLC on behalf of Society for Research in Child Development
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
All rights reserved. 0009-3920/2021/9206-0021
DOI: 10.1111/cdev.13623

fluctuations from study to study, male characters are consistently about twice as common as female characters in children's picture books, an effect largely driven by male rather than female authors (Hamilton, Anderson, Broaddus, & Young, 2006). Androcentrism remains evident in more recent books where one might imagine authors would be more alert to reducing bias (Lee & Chin, 2019; Moser & Masterson, 2013). Alongside frequency differences, male and female characters are depicted in different ways. Lewis et al., (2020) reported substantial gender stereotyping in the activities of main characters in 249 popular picture books written for children. When female characters are referenced, they are more likely to be in caretaking roles, whereas male characters lead more adventurous lives (Jackson & Gee, 2005; Macalister, 2011; Turner-Bowker, 1996). Gender stereotyping extends to nonhuman animal characters, and animals are much more likely to be given male names (McCabe et al., 2011), as are anthropomorphized inanimate objects, such as cars and machines (Berry & Wilkins, 2017).

Androcentrism is evident in language more broadly. Johns and Dye (2019) analyzed fiction and nonfiction written for adults as well as TV and film subtitles. They found a large and systematic preference for male names (around 59%–62%, depending on subcorpus used). This held across genres and was consistent over historical time, with contemporary text just as biased as older literature sampled across the last 200 years. The male bias was largely due to male authors tending to write more about male characters, whereas female authors use a more equal number of male and female names. Taken together, it is clear that written language talks differently about males and females and importantly for development, gender biases are apparent in picture books for preschoolers as well as curriculum books and readers for school-age children, and have been documented in different cultures and across different educational systems (Filipović, 2018; Lee & Chin, 2019). This prompts the question of how experience with language shapes the development of gender stereotypes and biases. As noted by Lewis and Lupyan (2020), gender associations in language might play a role in establishing cultural stereotypes, or language might simply reflect the stereotypes that already exist. While direct causal evidence is lacking, young children are sensitive to linguistic devices that influence stereotype formation (e.g., Cimpian & Markman, 2011; Rhodes, Leslie, Yee, & Saunders, 2019) and Lewis and Lupyan (2020) found strong relationships between gender associations in language and people's implicit

gender associations, an effect that held across 25 different languages.

Few studies have explored children's own writing. Romatowski and Trepanier-Street (1985) coded written language samples from 90 boys and 90 girls, spanning Grades 1–6. Male characters are featured more often in the children's writing. Another striking feature of their data was that boys wrote overwhelmingly about male characters—only 14% of characters introduced by boys were female. Girls were more balanced, with 44% of their characters being female (see also Many, 1989). These differences between boys and girls pattern like adult authors, with androcentrism largely attributable to male authors (Johns & Dye, 2019).

In summary, there is clear evidence of gender bias in written language. Analyses of children's literature suggest that androcentrism is present, holds over historical time and might be attributable to male authors. Lacking, however, is a large-scale analysis and full consideration of contemporary literature: existing studies are small and selective, especially once data are broken down by character and author gender. Study 1 addressed this by analyzing personal name usage in a large corpus of children's literature. Given the androcentrism reported in previous investigations, we predicted that male names would be used more often in children's books, particularly by male authors. In the Supporting Information, we also report several exploratory analyses enabled by the metadata embedded in the corpus (i.e., publication year, targeted age, and genre), along with a separate analysis on pronoun usage to test the generalizability of personal name effects.

In Study 2, we analyzed stories written by over 100,000 children. These were not written for the purpose of this investigation but like other naturally occurring data sets (Goldstone & Lupyan, 2016), they provide a window on underlying mental processes, uncontaminated by experimenter cues and bias. We predicted that children would produce more male names in their own writing. We also asked whether any gender bias would be driven by boys, as is the case for male adult authors. The availability of cross-sectional developmental data allowed us to test whether any biases change with age, and whether patterns are the same for boys and girls.

Study 1

This investigated personal name usage in children's books. We first obtained the distribution of male

and female names in the population, using birth registration data. We then analyzed a large developmental corpus of children's books to reveal any deviations and biases in personal name usage by male and female authors. Based on the previous literature, we predicted an overall male name bias, mainly due to male writers, and our analyses provided confirmatory tests of this hypothesis. Additional exploratory analyses investigated personal name usage by targeted age, genre, and year of publication, and these factors crossed with author gender. The results are detailed in the Supporting Information where we also report analyses on the frequency of male and female pronoun usage, complementing and replicating the main analyses on personal names. Data and analysis code are available at <http://osf.io/m9arj>.

Method

We obtained male and female names from birth registration data for 621,991 babies born in England and Wales in 2017, the latest data set available at the time of analysis (Office for National Statistics, 2017). About half were boys (322,902; 52%; girls 299,089; 48%). Names that occurred less

than three times were removed. Similar names with different spellings (e.g., John and Jon) were counted as separate names. Girl names varied more than boy names: of the 63,697 different names registered, 13,697 occurred more than three times and 55% were for girls, 45% for boys. Figure 1 shows that more boys share the most popular boy names than girls, perhaps reflecting the historical observation that girl names are more variable and girls may be more likely to receive androgynous names than boys (Liebersohn, Dumais, & Baumann, 2000).

Turning to personal name usage in children's literature, the reading part of the Oxford Children's Corpus comprises 34 million words taken from 13,154 documents written for children aged 5–13 years. It includes fiction, nonfiction, magazines, Web sites, and curriculum materials. We first extracted all words identified as proper nouns using the part of speech tagger NNP (indicating singular proper name) based on the Penn Treebank tagging system. These were then mapped to the names identified in the birth registration data. This method is not perfect; some names, like Alex, are assigned to both boys and girls for example. We used a cutoff of 60% to determine the primary gender for each

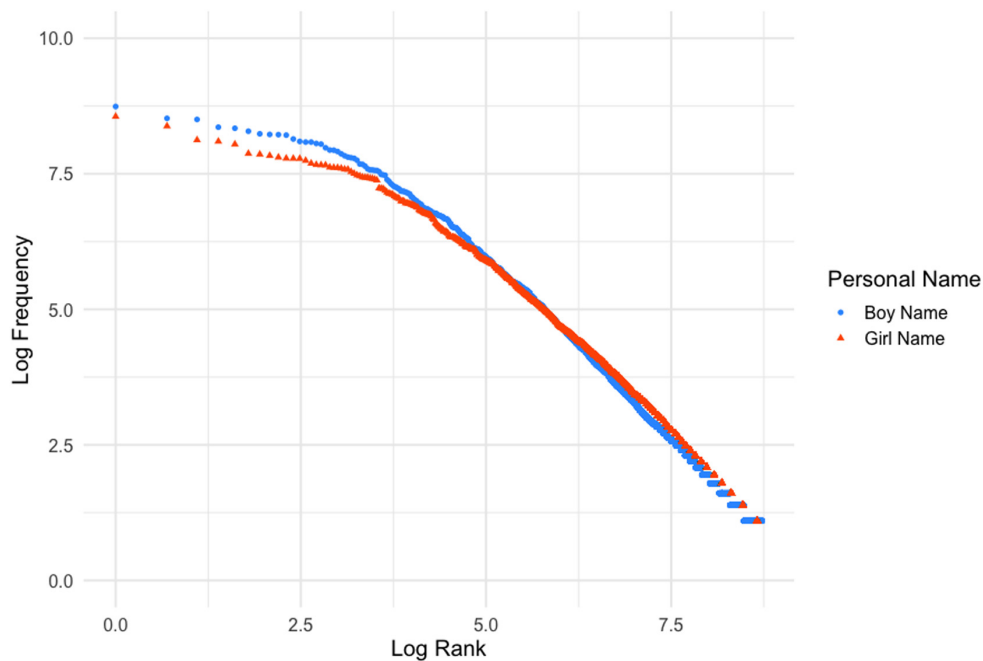


Figure 1. Zipf scale of boy versus girl names for babies born in England and Wales in 2017.

Note. A Zipf scale is a standard way to display frequency distributions of words (e.g., Johns & Dye, 2019; Piantadosi, 2014; Zipf, 1949). The *y*-axis shows the logarithm of the raw frequency of the name (where a higher number indicates more frequent occurrence) and the *x*-axis maps onto the logarithm of the name's rank (where a lower number indicates more frequent occurrence). This figure shows that high-frequency boy names are more frequent than high-frequency girl names.

name. That is, if a particular name was assigned to boys more than 60% of the time, we designated that name as a male name. Documents without information about title, author, genre, and year of publication were excluded. Most of the documents removed were encyclopedia entries, Web sites, and some curriculum materials. The remaining corpus (729 books containing over 24 million words) comprised our data set.

Results and Discussion

The corpus contained 4,028 unique personal names that occurred 630,400 times in total. As summarized in Figure 2, male names occurred more often (62%; $N = 392,588$; female 38%; $N = 237,812$). This is significantly higher than the 52% seen in the population baseline, $\chi^2(1) = 26,670$ $p < .001$. There were also more types of male name, with 2,142 different names (53%) compared with 1,886 female names (47%). This is counter to the statistics in birth registration data where type frequency was higher for female names (55%) than male names (45%), $\chi^2(1) = 108.84$, $p < .001$.

To examine whether author gender influenced the likelihood of using more male names, we first identified the author's gender of each document in

the corpus by mapping the author's first name to the birth registration data. We used manual coding for names that did not have a match (e.g., unpacking names involving abbreviation) or when authorship consisted of multiple authors of more than one gender (multiple authorship within the same gender was categorized as one). Some documents were impossible to code (e.g., when individual author names were not specified) and these were excluded. In the remaining texts, personal names occurred 357,069 times; 45% (159,579) were produced by male authors, 54% (193,132) by female authors, and 1% (4,358) were co-authored by authors of more than one gender (statistics were entered in models (see Supporting Information) but will not be discussed below due to sparse data).

For male name usage, the gender of the author did not matter: 51% (112,998) of occurrences were by male authors, 48% (105,767) by female authors, and 1% (1,655) by co-authors. This contrasts sharply with the pattern for female names where 34% (46,581) were produced by male authors, 64% (87,365) by female authors, and 2% (2,703) by co-authors. Splitting the data by author gender, of the 159,579 names written by male authors, 71% (112,998) were male names, much higher than the 52% expected from baseline, $\chi^2(1) = 22,621$,

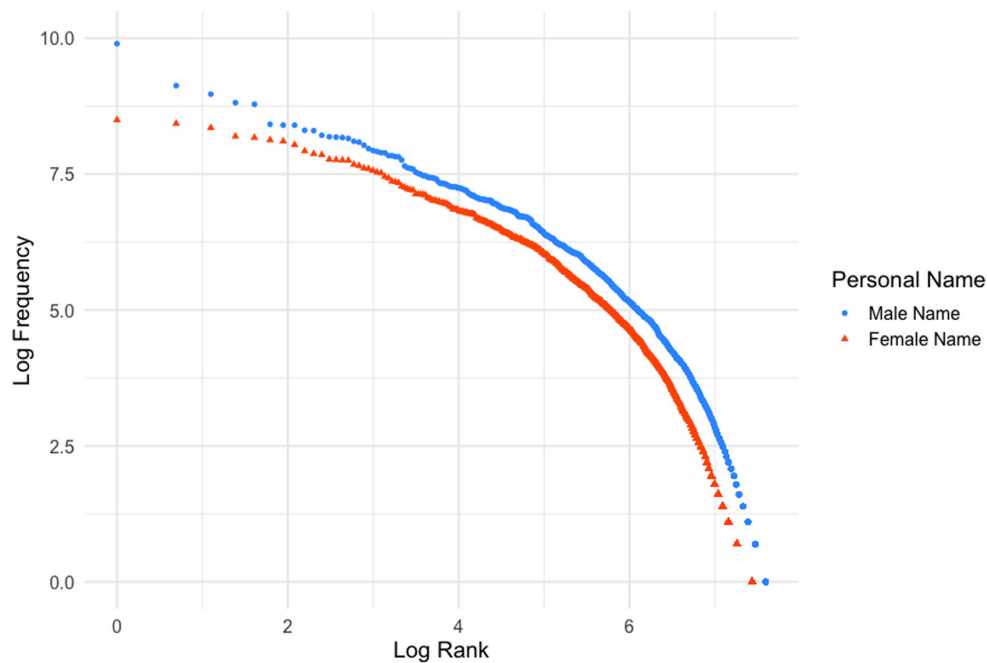


Figure 2. Zipf scale of male versus female names in Oxford Children's Corpus (texts written for children).

$p < .001$. Of the 193,132 names written by female authors, 55% (105,767) were male, also above baseline, $\chi^2(1) = 591.18$, $p < .001$.

These analyses, visualized in Figure 3, reveal a striking bias as a function of author gender. Clearly, male authors show a strong and consistent preference to write about males rather than females, relative to female authors. We used a generalized linear mixed effects model (glmer) to investigate the effect of author gender (independent variable) on the gender of the personal names used (dependent variable); this can accommodate unbalanced sample size across groups (Brauer & Curtin, 2018). Document served as a random variable, given documents might differ in terms of the nature of names included. Random slopes were not included as author gender does not vary within a document. The model produced a significant effect of author gender: when female authors were treated as the reference group, male authors showed significantly higher male name usage ($b = 0.80$, $SE = .10$, $z = 8.37$, $p < .001$).

Exploratory analyses (Supporting Information) showed that the male names dominated in the same way in both fiction and nonfiction, across historical time, and are present regardless of targeted age. It also extends to gendered pronouns (e.g., she, her, hers, herself; he, his, him, himself).

In summary, Study 1 found no gender imbalance in the general population according to birth registration data but a clear dominance of male names in children's books. Male authors were largely responsible for this increased use of male names. This replicates the overrepresentation of male names in adult books (Johns & Dye, 2019) and echoes earlier findings from small-scale investigations of children's books (Hamilton et al., 2006; Weitzman et al., 1972).

Study 2

The writing component of the Oxford Children's Corpus contains stories written by 5- to 13-year-old children across the United Kingdom for an annual national writing competition hosted by BBC Radio 2 in collaboration with Oxford University Press. Children may write on any topic they wish—the only constraint is that they must use no more than 500 words. We analyzed the 105,369 entries submitted in 2019, totaling over 46 million words. Following Study 1, we extracted names from the corpus and matched them with birth registration data to obtain gender. We then tested our hypotheses with confirmatory analyses addressing whether children show personal name bias in their writing, whether

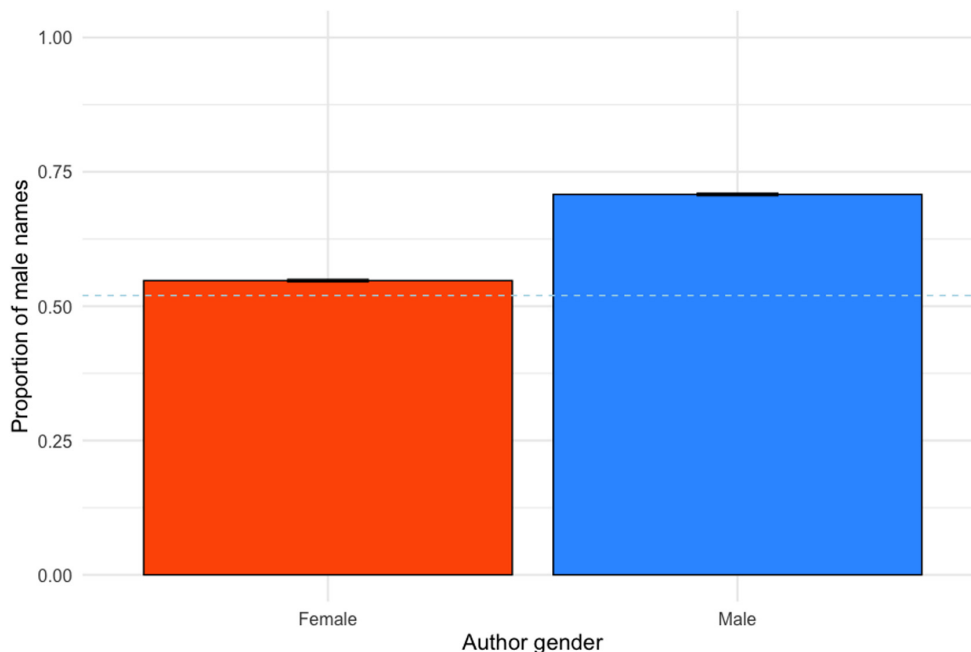


Figure 3. The proportion of male names by author's gender in Oxford Children's Corpus. Note. The dotted line indicates the proportion of male names in the birth data (52%).

boy and girl authors differ, and whether there are differences by age.

Results and Discussion

Figure 4 shows the distribution of male and female names in children's writing. Male names occurred more often with 654,320 (56%) tokens (cf. 513,579 (44%) female names). The proportion of male names was high relative to birth data, $\chi^2(1) = 7581.9$ $p < .001$. Although the proportion of unique male names was 48%, this was higher than expected birth data, $\chi^2(1) = 15.59$, $p < .001$. Thus, the children used a slightly wider range of female names than male names, but they used them less often. For male names, 363,189 (56%) were produced by boys and 291,131 (45%) by girls. The pattern of data for female names is different: only 66,615 (13%) female names were written by boys in contrast to the 446,964 (87%) produced by girls.

Figure 5 shows the data split by author gender (this was embedded in the corpus as metadata for each story). Boys produced 429,804 names. Strikingly, only 15% ($N = 66,615$) were female—the remaining 85% ($N = 363,189$) depicted male characters, much greater than expected from baseline, $\chi^2(1) = 181,895$, $p < .001$. Fifty-nine percent of all unique male names produced by boys were male,

higher than female names ($N = 2,168$; 59%, $N = 1,532$; 41%, respectively), again higher than the 45% in the birth data, $\chi^2(1) = 276.29$, $p < .001$. In contrast, girls produced 31% ($N = 291,095$) male names compared to 61% ($N = 446,964$) female names, less than expected by the baseline, $\chi^2(1) = 46,643$, $p < .001$. Girls also used fewer unique male names ($N = 1,914$, 43%) than female names ($N = 2,535$, 57%), showing that girls also write more about their own gender, but to a lesser extent than boys. Confirming these observations, a glmer model with author gender as the independent variable and gender of the personal name as the dependent variable confirmed that boys produced significantly more male names than girls ($b = 3.48$, $SE = .02$, $z = 167.63$, $p < .001$).

Adding author age as a continuous fixed factor (centered and scaled) to the model showed that older children used more male names than younger children ($b = 0.01$, $SE = .001$, $z = 20.73$, $p < .001$). The main effect of author gender remained strong, with boys producing more male names than girls ($b = 0.66$, $SE = .01$, $z = 48.46$, $p < .001$). Author gender also interacted with age. As shown in Figure 6, boys showed a constant male name preference across age, whereas girls became more likely to use male names with age ($b = -0.03$, $SE = .001$, $z = -21.73$, $p < .001$). Although older girls produced more male names than younger girls, they

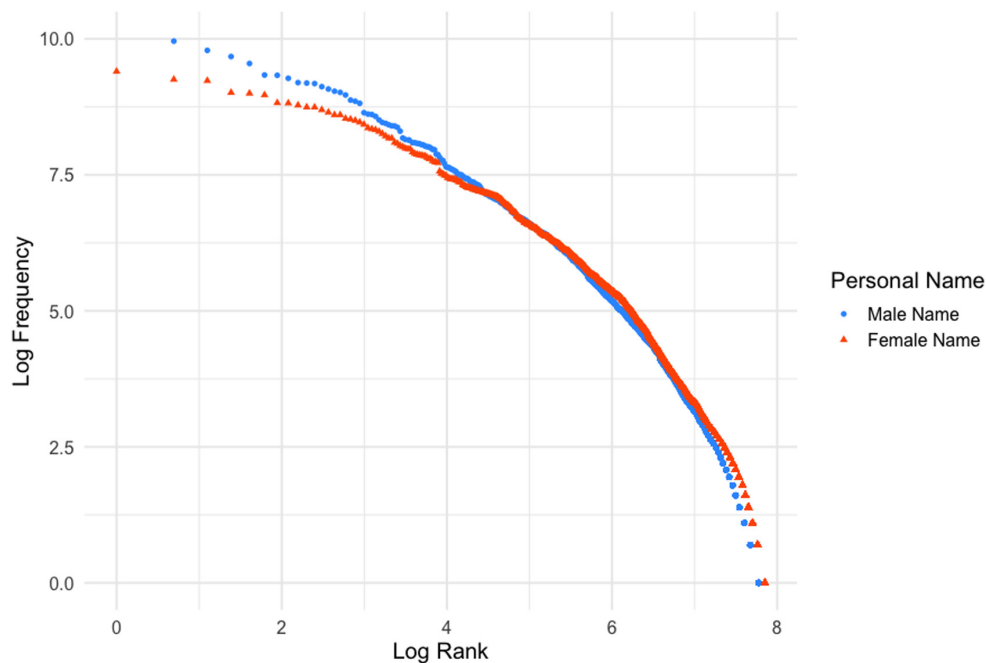


Figure 4. Zipf scale of male versus female names in children's own writing.

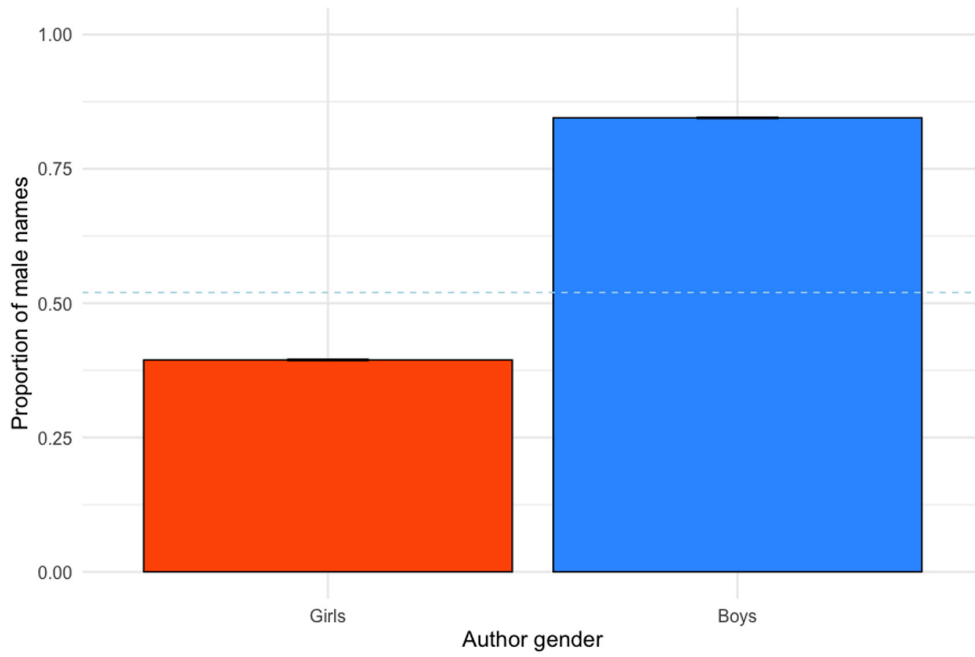


Figure 5. Proportion of male names in writing by boys and girls.
 Note. The dotted line indicates the proportion of male names in the birth data (52%).

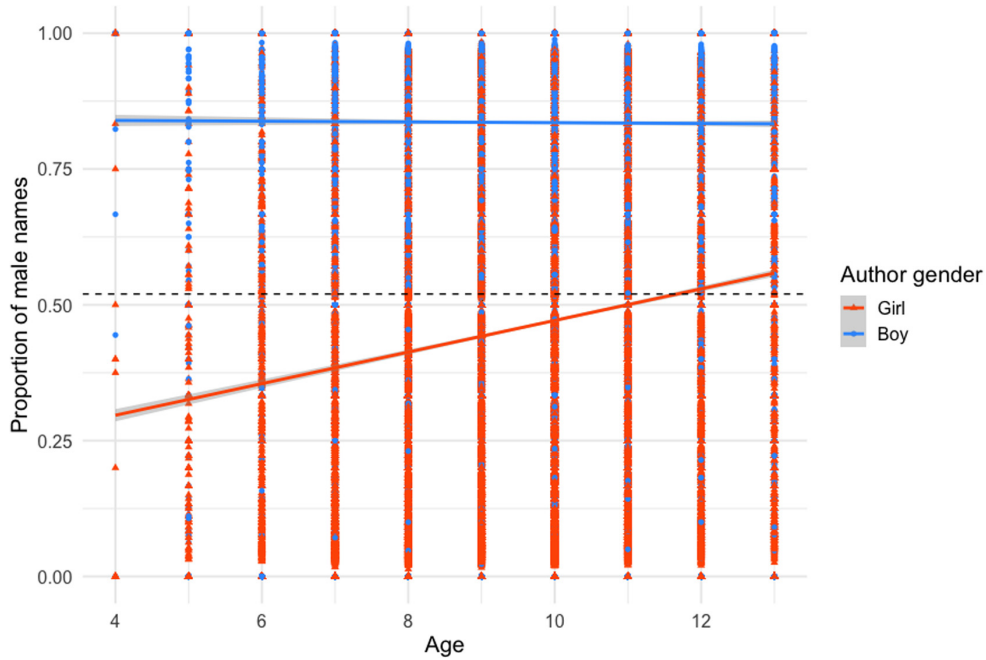


Figure 6. Proportion of male names in the writing by boys and girls across age.
 Note. The gray dotted line indicates the proportion of male names in the birth data (52%).

still used far fewer male names than boys of the same age. With age, girls came to produce approximately 50% male names, whereas boys averaged 85% male names.

General Discussion

We used large developmental corpora to investigate the usage of male and female names in children's written language and found clear evidence of androcentrism. Like language written for adults (Johns & Dye, 2019), children's books contain more male names, largely due to male authors. Children themselves also used more male names than female names when writing and here there were striking differences for boys and girls, and developmental differences that interacted with author gender. Boys rarely wrote about female characters, whereas girls were more balanced, mirroring the pattern seen for adult female authors in Study 1. We also found that younger children tended to write more about characters of their own gender and while boys continued this trend, girls became less egocentric as they got older. The older girls produced approximately 50% male names, whereas the older boys averaged 85% male names. The finding that children initially produce more character names of their own gender is consistent with findings from the gender schema literature where it is well established that one's own gender shapes how information is organized and remembered (Bem, 1981). Children are sensitive to gender schema from a young age. For example, stereotypically feminine activities and female characters are better remembered by girls, whereas activities and characters associated with males are better remembered by boys (Signorella, Bigler, & Liben, 1997).

Our findings echo those from the developmental gender cognition literature, including the Draw-a-Scientist task. Miller, Nolla, Eagly, and Uttal (2018) reported that drawings become more male through childhood, largely driven by girls shifting from drawing female scientists to producing more males; boys, in contrast, showed a stable rate of in-group bias across development. The gender-brilliance paradigm asks children whether their own gender is "really, really smart" and by 6 years of age, girls are less likely to believe so than boys. Girls are also less interested in participating in activities said to be for brilliant children (Bian, Leslie, & Cimpian, 2017; Storage, Charlesworth, Banaji, & Cimpian, 2020), or associated with masculine cues like video games (Cheryan, Davies, Plaut, & Steele, 2009) or less family-friendly and community-motivated

(Diekman, Weisgram, & Belanger, 2015). These cultural stereotypes have been linked with particular outcomes such as men having desirable traits for intellectual careers and overtime, they may well become reflected in language usage. On this view, children's growing experience with the world shapes their perception and embodiment of cultural stereotypes and this is then reflected in the language they produce. Lewis and Lupyan (2020) remind us that a causal relationship might also operate in reverse such that exposure to gender associations in language shapes the development of cultural stereotypes. Here, systematic biases in language input will become reflected in children's cognition, and their language output. Presumably, these two views are not mutually exclusive and could be reciprocally connected (Bem, 1981).

Building on the idea that language experience affects gender perception, one explanation for why boys and girls show different patterns of androcentrism in their writing is that boys tend to read less than girls, and lag behind in reading achievement (Chiu & McBride-Chang, 2006). Since children's books contain more description of males, girls gradually produce more writing about men as they read more and therefore experience more androcentrism. For boys, there is less pressure from language experience to shift from their initial in-group bias and therefore their tendency to use male names maintains across development. At the same time, because boys read less, they also encounter fewer female characters in books than girl peers. Additionally, there are systematic differences in the types of things boys and girls choose to read and plausibly, these might deliver different biases in written language input. Surveying over 8,000 six- to sixteen-year-olds, Clark and Foster (2005) found that while adventure, comedy, and horror/ghost stories were popular with boys and girls, girls were more likely to enjoy romance/relationships and books about animals; boys reported stronger preferences for science fiction/fantasy, war/spy-related, crime/detective and books about sport. Similarly, there is evidence of gender bias in what caregivers choose to read to preschool children with girls more likely to hear stereotypically female content (Lewis et al., 2020) and McGeown (2015) found that identification with female traits predicted motivation to read in a group of 9- to 10-year-olds, and the likelihood of choosing to read female-oriented fiction.

Whether these genres of children's fiction differ in the strength of male name bias is not known, but analysis of adult fiction offers relevant evidence. Johns and Dye (2019) sampled ten genres (fantasy,

historical, mystery, romance, thriller, young adult, crime, literature, horror, and science fiction) and found that personal names were more likely to be male in each. However, the bias was smallest in romance, young adult, and historical fiction and largest in science fiction and horror. Notably, variation in the male name bias across genres aligned with author gender; as Johns and Dye (p. 1609) note:

“When a genre has a greater degree of male authorship, there is a corresponding increase in the frequency of male names. When a genre is mainly written by female authors, there is a much more egalitarian use of male and female names.”

Given these findings, it is reasonable to predict that boys encounter fewer female names when reading as they read less overall, and when they do read, they are more likely to be choosing material that contains a higher proportion of male names. Boys might sample more male bias in their reading experience than girls, and this might play out in their own writing such that boys continue to show in-group bias in personal name usage, whereas girls become more egalitarian.

As Caliskan, Bryson, and Narayanan (2017, p. 183) comment in the context of machine learning, text corpora “contain recoverable and accurate imprints of our historic biases.” While language might reflect cultural biases, the capacity for language to create and sustain stereotypes should not be ignored (Lewis & Lupyan, 2020). Language is important to children’s beliefs about gender stereotypes (e.g., Cimpian & Markman, 2011; Liben, Bigler, & Krogh, 2002; Rhodes et al., 2019) and once children can read, written language provides a powerful substrate for learning about the world: young adults typically read millions of words per year (Brysbaert, Stevens, Mander, & Keuleers, 2016). Although our findings cannot confirm a causal link between children’s exposure to written language and their writing, we show evidence of similar androcentrism across these two domains. Studies are needed to confirm the nature of the relationship by experimentally manipulating children’s reading experience and examining subsequent changes in their gender representation.

References

- Bailey, A., & LaFrance, M. (2017). Who counts as human? Antecedents to androcentric behavior. *Sex Roles, 76*, 682–693. <https://doi.org/10.1007/s11199-016-0648-4>
- Bailey, A., LaFrance, M., & Dovidio, J. F. (2019). Is man the measure of all things? A social cognitive account of androcentrism. *Personality and Social Psychology Review, 23*, 307–331. <https://doi.org/10.1177/1088868318782848>
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*, 354–364. <https://doi.org/10.1037/0033-295X.88.4.354>
- Berry, T., & Wilkins, J. (2017). The gendered portrayal of inanimate characters in children’s books. *Journal of Children’s Literature, 43*, 4–15.
- Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children’s interests. *Science, 355*, 389–391. <https://doi.org/10.1126/science.aah6524>
- Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*, 389–411. <https://doi.org/10.1037/met0000159>
- Brysbaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant’s age. *Frontiers in Psychology, 7*, 1–11. <https://doi.org/10.3389/fpsyg.2016.01116>
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*, 183–186. <https://doi.org/10.1126/science.aal4230>
- Cheryan, S., Davies, P. G., Plaut, V. C., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*, 1045–1060. <https://doi.org/10.1037/a0016239>
- Chiu, M. M., & McBride-Chang, C. (2006). Gender, context, and reading: A comparison of students in 43 countries. *Scientific Studies of Reading, 10*, 331–362. https://doi.org/10.1207/s1532799xssr1004_1
- Cimpian, A., & Markman, E. M. (2011). The generic/non-generic distinction influences how children interpret new information about social others. *Child Development, 82*, 471–492. <https://doi.org/10.1111/j.1467-8624.2010.01525.x>
- Clark, C., & Foster, A. (2005). *Children’s and young people’s reading habits and preferences the who, what, why, where and when*. London, UK: National Literacy Trust.
- Diekmann, A. B., Weisgram, E. S., & Belanger, A. L. (2015). New routes to recruiting and retaining women in STEM: Policy implications of a communal goal congruity perspective. *Social Issues and Policy Review, 9*, 52–88. <https://doi.org/10.1111/sipr.12010>
- Filipović, K. (2018). Gender representation in children’s books: Case of an early childhood setting. *Journal of Research in Childhood Education, 32*, 310–325. <https://doi.org/10.1080/02568543.2018.1464086>
- Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring

- data sets. *Topics in Cognitive Science*, 8, 548–568. <https://doi.org/10.1111/tops.12212>
- Hamilton, M. C., Anderson, D., Broaddus, M., & Young, K. (2006). Gender stereotyping and under-representation of female characters in 200 popular children's picture books: A twenty-first century update. *Sex Roles*, 55, 757–765. <https://doi.org/10.1007/s11199-006-9128-6>
- Hopkins, E. J., & Weisberg, D. S. (2017). The youngest readers' dilemma: A review of children's learning from fictional sources. *Developmental Review*, 43, 48–70. <https://doi.org/10.1016/j.dr.2016.11.001>
- Hyde, J. S. (1984). Children's understanding of sexist language. *Developmental Psychology*, 20, 697–706. <https://doi.org/10.1037/0012-1649.20.4.697>
- Jackson, S., & Gee, S. (2005). 'Look Janet', 'No you look John': Constructions of gender in early school reader illustrations across 50 years. *Gender and Education*, 17, 115–128. <https://doi.org/10.1080/0954025042000301410>
- Johns, B. T., & Dye, M. (2019). Gender bias at scale: Evidence from the usage of personal names. *Behavior Research Methods*, 51, 1601–1618. <https://doi.org/10.3758/s13428-019-01234-0>
- Lee, J., & Chin, A. (2019). Are females and males equitably represented? A study of early readers. *Linguistics and Education*, 49, 52–61. <https://doi.org/10.1016/j.linged.2018.12.003>
- Lewis, M., Borkenhagen, M. C., Converse, E., Lupyan, G., & Seidenberg, M. S. (2020). *What might books be teaching young children about gender?* Manuscript submitted for publication. <https://doi.org/10.31234/osf.io/ntgfe>
- Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, 4, 1021–1028. <https://doi.org/10.1038/s41562-020-0918-6>
- Liben, L. S., Bigler, R. S., & Krogh, H. R. (2002). Language at work: Children's gendered interpretations of occupational titles. *Child Development*, 73, 810–828. <https://doi.org/10.1111/1467-8624.00440>
- Lieberson, S., Dumais, S., & Baumann, S. (2000). The instability of androgynous names: The symbolic maintenance of gender boundaries. *American Journal of Sociology*, 105, 1249–1287. <https://doi.org/10.1086/210431>
- Logan, J. A. R., Justice, L. M., Yumus, M., & Chaparro-Moreno, L. J. (2019). When children are not read to at home: The million word gap. *Journal of Developmental & Behavioral Pediatrics*, 40, 383–386. <https://doi.org/10.1097/DBP.0000000000000657>
- Macalister, J. (2011). Flower-girl and bugler-boy no more: Changing gender representation in writing for children. *Corpora*, 6, 25–44. <https://doi.org/10.3366/cor.2011.0003>
- Many, J. E. (1989). Sex roles from a child's point of view: An analysis of children's writing. *Reading Psychology*, 10, 357–370. <https://doi.org/10.1080/0270271890100405>
- McCabe, J., Fairchild, E., Grauerholz, L., Pescosolido, B. A., & Tope, D. (2011). Gender in twentieth-century children's books: Patterns of disparity in titles and central characters. *Gender & Society*, 25, 197–226. <https://doi.org/10.1177/0891243211398358>
- McGeown, S. P. (2015). Sex or gender identity? Understanding children's reading choices and motivation. *Journal of Research in Reading*, 38, 35–46. <https://doi.org/10.1111/j.1467-9817.2012.01546.x>
- Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of U.S. Draw-a-Scientist studies. *Child Development*, 89, 1943–1955. <https://doi.org/10.1111/cdev.13039>
- Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear : Picture books and the statistics for language learning. *Psychological Science*, 26, 1489–1496. <https://doi.org/10.1177/0956797615594361>
- Moser, F., & Masterson, J. (2013). Are there signs of change in gendered language use in children's early reading vocabulary? *Gender and Language*, 8, 71–90. <https://doi.org/10.1558/genl.v8i1.71>
- Office for National Statistics. (2017). *Baby names in England and Wales: 2017*. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/babynamesenglandandwales/2017>
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21, 1112–1130. <https://doi.org/10.3758/s13423-014-0585-6>
- Rhodes, M., Leslie, S.-J., Yee, K. M., & Saunders, K. (2019). Subtle linguistic cues increase girls' engagement in science. *Psychological Science*, 30, 455–466. <https://doi.org/10.1177/0956797618823670>
- Romatowski, J., & Trepanier-Street, M. (1985). Attributes and roles assigned to characters in children's writing: Sex differences and sex-role perceptions. *Sex Roles*, 13, 263–272. <https://doi.org/10.1007/BF00288084>
- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1997). A meta-analysis of children's memories for own-sex and other-sex information. *Journal of Applied Developmental Psychology*, 18, 429–445. [https://doi.org/10.1016/S0193-3973\(97\)80009-3](https://doi.org/10.1016/S0193-3973(97)80009-3)
- Storage, D., Charlesworth, T. E. S., Banaji, M. R., & Cimpian, A. (2020). Adults and children implicitly associate brilliance with men more than women. *Journal of Experimental Social Psychology*, 90. <https://doi.org/10.1016/j.jesp.2020.104020>
- Turner-Bowker, D. M. (1996). Gender stereotyped descriptors in children's picture books: Does "curious Jane" exist in the literature? *Sex Roles*, 35, 461–488. <https://doi.org/10.1007/BF01544132>

- Weitzman, L. J., Eifler, D., Hokada, E., & Ross, C. (1972). Sex-role socialization in picture books for pre-school children. *American Journal of Sociology*, 77, 1125–1150. Retrieved from <https://doi.org/10.1086/225261>
- Young, T. A., Moss, B., & Cornwell, L. (2007). The Classroom library: A place for nonfiction. *Nonfiction in Its Place*, 48. Retrieved from https://scholarworks.wmich.edu/reading_horizons/vol48/iss1/3
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Oxford, UK: Addison-Wesley Press.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Results of exploratory analyses, including male name bias across genre, targeted age, and time period in children's books, as well as pronoun use in both children's books and children's own writing.