Full Length Article

Gender associations in the twentieth-century English-language literature

Daniel Schulz [a], Štěpán Bahník [b],*

[a] Department of Psychology, University of Würzburg, Röntgenring 10, Würzburg 97070, Germany
[b] Faculty of Business Administration, University of Economics, Prague, náměstí Winstona Churchilla 4, Prague 130 67, Czech Republic

ABSTRACT

Gender associations may be partly learned from print media, including literature. Using Google Books Ngram corpus, we explored the depiction of male and female characters in the twentieth-century English-language fiction. By analyzing adjective-noun bigrams, we examined adjectives used in association with "man", "woman", "boy", and "girl". Men were described in more positive terms than women. Girls were depicted more positively than boys at the beginning of the twentieth century, but the tendency reversed in the middle of the century. Boys were described in more masculine terms than girls; however, men were described in similarly masculine adjectives as women. Despite limitations of interpretability of the results, the study presents a possible approach of exploring past characterization of the two genders.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

People have different associations with men and women, which form a basis of their gender stereotypes. Even though stereotypes differ between various gender subgroups (Fiske, Xu, Cuddy, & Glick, 1999), men tend to be generally seen as more competent and women as more warm (Kite, Deaux, & Haines, 2008). Women have also been found to be perceived generally more favorably than men (Eagly, Mladinic, & Otto, 1991). Given that gender stereotypes tend to be accurate (Swim, 1994), people may learn them through observation and induction. They are also transmitted by social learning and communicated by family, peers, and media (Kite et al., 2008).

Studies examining the depiction of men and women in media have shown that gender stereotypes are mirrored in various forms of media. A recent meta-analysis of studies of gender depiction in advertisements showed that women are more likely to be portrayed at home, associated with domestic products, not speaking, and in a dependent role (Eisend, 2010). In children's cartoons, male characters were more likely to be portrayed as independent, assertive, athletic, and technical, while female character were more likely to be portrayed as emotional, warm, affectionate, and frail (Thompson & Zerbinos, 1995). Similarly, male characters in popular movies were more likely than female characters to have a lead-

ership role, occupational power over others, and goals (Lauzen & Dozier, 2005). In an analysis of adjectives used to describe characters in prize-winning children's books from 1984 to 1994, Turner-Bowker (1996) found that the adjectives used for male characters were evaluated as more masculine, active, and potent and the adjectives used to describe female characters were evaluated more positively.

Apart from the difference in depiction of male and female characters in various media, the proportion of characters of the two genders also differs. Even though the ratio of male to female characters in children's books varied over the twentieth century, the characters were more likely to be male than female for most of the century. Only in 1910s and 1990s was the proportion of male and female characters approximately equal (McCabe, Fairchild, Grauerholz, Pescosolido, & Tope, 2011). Women have also been found to be underrepresented among television characters (Elasmar, Hasegawa, & Brain, 1999), in movies (Lauzen & Dozier, 2005), in fiction books (Underwood, Bamman, & Lee, 2018), and in children's cartoons (Thompson & Zerbinos, 1995).

Recent creation of the Google Books Ngram corpus (Michel et al., 2011) has enabled the study of depiction of male and female characters in a large volume of the fiction literature. Google Books Ngram corpus contains data about word usage from more than 5 million books. The data relate to *n-grams*, which are sequences of *n* words used in a text. Previous studies used the corpus to study, for example, the use of emotion-related words in books (Acerbi, Lampos, Garnett, & Bentley, 2013), the use of individualistic words

and phrases (Twenge, Campbell, & Gentile, 2012a), or censorship of topics and regularization of irregular verbs (Michel et al., 2011). Some studies have also explored issues related to gender. Mason, Kuntz, and McGill (2015) found that women appeared in English-language books less often than men and that, in comparison to women, men tended to be relatively more often described as "young" than "old". Twenge, Campbell, and Gentile (2012b) found that male pronouns occurred more often in U.S. books than female pronouns and that the ratio of their use was highest during the 1950s and early 1960s, and that it has been decreasing from then on. Finally, Ye, Cai, Chen, Wan, and Qian (2018), who used a similar method as the present study, found that in English-language books, men tended to be more often described in terms associated with agreeableness, extraversion, conscientiousness, and neuroticism traits of the Big Five model and that they were described similarly often as women in terms associated with openness. However, Ye et al. (2018) did not take into account that men were generally relatively more often described with an adjective than women, which could have confounded some of their results. That is, men could be more often described by adjectives associated with a certain factor just because they are more often described by any adjective.

In the present study, we use the Google Books Ngram corpus to explore the depiction of men and women in the English-language literature and its development in the 20th century. In particular, we focus on adjectives that were used in association with nouns "man", "woman", "boy", and "girl". The four nouns were selected to study both the effect of gender and a possible difference between portrayal of adults and children. We explored whether male or female characters were depicted in more masculine and positive terms. We also examined how these differences, for both adult and children characters, changed during the century. Relatedly, we examined the variability of the terms used to depict the two genders. High variability of positivity could indicate, for example, that the characters are described in black-and-white terms, while low variability could suggest more nuance in the depiction. Finally, we looked at similarities of terms used in association with the selected nouns and their development during the century. The data on historical development of the perception of men and women are scarce and difficult to obtain retrospectively. Insofar that the depiction of men and women in the literature reflects their contemporary perception, the method used in the present study could allow us to study views of men and women indirectly. More importantly, the depiction of men and women in books is of interest in itself given that it might influence opinions and views of readers (Gelman, 2009; Lewis & Lupyan, 2019; Lupyan & Lewis, 2017; Paluck & Green, 2009).

2. Method[1]

2.1. Bigrams

We searched for adjective-noun bigrams in the Google Books Ngram data (http://storage.googleapis.com/books/ngrams/books/-datasetsv2.html). In particular, we used the English fiction subset of the dataset (version 20120701). We used syntactic annotations (Lin, Michel, Aiden, Orwant, Brockman, & Petrov, 2012) to select only bigrams in which adjectives (described below) were related to the nouns we studied — "man", "woman", "boy", and "girl" (as well as "person" and "child", which were however used only for the analysis of similarity of descriptions). The selected nouns were chosen to represent most frequent nouns used for both genders and for both adults and children. The Google Books Ngram corpus

includes only n-grams that occurred at least 40 times across the corpus, which means that some n-grams for less frequent nouns might have been missing. Other possible nouns (e.g., "husband" and "wife") refer to specific roles rather than to general people of the two genders, which we were interested in, and most of them would also not be applicable to both adults and children.

We used only the data from years between 1900 and 2000. The precise cutoff was arbitrary, but the years preceding 1900 tended to have insufficient amount of data and after 2000 the sampling of books changed (http://www.culturomics.org/Resources/faq#-dataquality). The books included in the database published after 2000 no longer came predominantly from libraries, so their inclusion could have led to spurious changes in trends after 2000, which would reflect only the differences in sampling rather than real changes in depiction of male and female characters.

2.2. Adjectives

We used a list of adjectives describing people's traits collected by Anderson (1968). In his study, Anderson used 18,000 trait names compiled by Allport and Odbert (1936) and reduced them to a list of 555 words by including only appropriate words used frequently for describing other people. To further reduce this list of 555 trait words, we selected 100 adjectives that were most often used in books in combination with each of the four studied nouns between years 1900 and 2008.[2] Since the adjectives did not completely overlap, we ended up with 155 adjectives. The full list of adjectives can be found at: https://osf.io/mrj5e/

To make general trends easier to interpret, we characterized the adjectives in terms of several dimensions: competence, warmth, desirability, and masculinity. Competence and warmth are considered universal dimensions of social cognition (Fiske, Cuddy, & Glick, 2007). Desirability reflects valence of the traits and allows us to compare our ratings of adjectives with ratings obtained half a century ago by Anderson (1968). Finally, the dimension of masculinity allows us to examine whether changing sex roles might have resulted in female characters being described in increasingly masculine terms and male characters in increasingly feminine terms during the century (cf. Diekman & Eagly, 2000).

Three hundred thirty-nine US Amazon Mechanical Turk workers[3] rated the adjectives on a randomly chosen dimension out of masculinity, competence, desirability, and warmth. The adjectives were rated on a 7-point scale adapted from Anderson (1968) in terms of how much they are associated with femininity (0) or masculinity (6), incompetence (0) or competence (6), and coldness (0) or warmth (6), or how much they are undesirable (0) or desirable (6).[4] The evaluation of the adjectives in terms of their desirability was stable between the study by Anderson (1968) and our survey (see Table 1). The stability of ratings over 49 years suggests that the adjectives kept similar connotations over the time. The adjective ratings for dimensions desirability, competence, and warmth correlated strongly between each other (see Table 1), so we computed their averages and used the composite score in analysis

---

[1] Data, analysis code, and the program used to download and extract raw data can be found on: https://osf.io/egpr5/. The study was not pre-registered.

[2] We based the selection on the range ending with 2008 because we originally intended to include years up to 2008 in the study and we did not obtain ratings for the few adjectives that would have been selected if we based the selection on the range ending in 2000.

[3] Data from participants who failed to answer at least one of two control items instructing them to pick a specific rating ($n = 64$) and from participants who finished the survey in less than 200 s or did not answer all questions ($n = 66$) were excluded from analysis. The sample size was deemed adequate to obtain average ratings of adjectives that would be sufficiently precise for the purpose of the study. Consistently, the average standard error of the mean (SE) of the ratings was 0.126 and the maximum SE was 0.210.

[4] We actually used a scale from 1 to 7, but we subtracted 1 from the ratings to make the results comparable to Anderson (1968) who used the scale from "least favorable or desirable" (0) to "most favorable or desirable" (6).

Table 1
Correlations between the dimensions used for adjective ratings. The table shows correlations between likeability ratings from Anderson (1968; 1), the composite positivity score (2) computed from ratings on dimensions 4–6, and adjective ratings on four dimensions obtained in the present study (3–6).

Variable	M	SD	1	2	3	4	5
1. Likeability (1968)	3.50	1.41					
2. Positivity	3.40	1.37	0.95**				
3. Masculinity	2.97	0.77	−0.28**	−0.25**			
4. Desirability	3.49	1.78	0.95**	0.99**	−0.22**		
5. Competence	3.41	1.30	0.89**	0.95**	−0.10	0.95**	
6. Warmth	3.30	1.20	0.89**	0.92**	−0.42**	0.89**	0.75**

** $p < .01$.

(the dimension called "positivity"; $\alpha = 0.94$). A positive correlation of traits in terms of their association with competence and warmth has been observed before and it has been shown to be due to the shared association with positive valence (Suitner & Maass, 2008).

2.3. Measures

The extraction yielded in total 5,795,481 syntactic bigrams over the four nouns combined. To study the description of men, women, boys, and girls in terms of positivity and masculinity, we used composite scores computed for each combination of a noun, dimension, and year (e.g., the positivity of adjectives used for depiction of "man" in the year 1950). To obtain the composite scores, we computed a weighted average of adjective ratings. That is, we multiplied average ratings on the dimension of the 100 adjectives most frequently associated with the noun with the proportion of bigrams from the given year that contained the adjective-noun bigram. For the year y, dimension d, and noun n, the composite score CS was thus computed as:

$$CS_{y,d,n} = \frac{\sum_{a \in A_n} f_{a,y,d,n} \times AR_{a,d}}{\sum_{a \in A_n} f_{a,y,d,n}}$$

where f is frequency, a is an adjective, A is the set of the most frequent adjectives, and AR is the average rating of an adjective. We used these composite scores in subsequent analyses.

Similarly, weighted standard deviation WSD, indicating variability of adjectives in terms of dimension d used to describe a noun n in a year y, was computed as:

$$WSD_{y,d,n} = \sqrt{\frac{\sum_{a \in A_n} f_{a,y,d,n} \times (AR_{a,d} - CS_{y,d,n})^2}{\sum_{a \in A_n} f_{a,y,d,n}}}$$

To find out how much an adjective a influenced differences between composite scores of nouns n_1 and n_2, we computed influence scores IS as:

$$IS_{a,d,n_1,n_2} = \frac{\sum_{y \in Y} CS'_{y,d,n_1} - \sum_{y \in Y} CS'_{y,d,n_2}}{l_Y}$$
$$- \frac{\sum_{y \in Y} CS''_{y,d,n_1} - \sum_{y \in Y} CS''_{y,d,n_2}}{l_Y}$$

where Y is the set of all years (1900–2000), l_Y is the number of years (i.e., 101 in our case), CS' is a composite score computed for the set of adjectives A', and CS'' is a composite score computed for the set of adjectives A'', where:

$$A' = A_{n_1} \cup A_{n_2} \text{ and } A'' = A_{n_1} \cup A_{n_2} - \{a\}.$$

The influence score for an adjective was thus computed by subtracting differences in composite scores for two nouns with and without the adjective. That is, the influence score shows how much would the average difference in composite scores of two nouns for a given dimension change if the adjective was not included in computation of the composite scores.
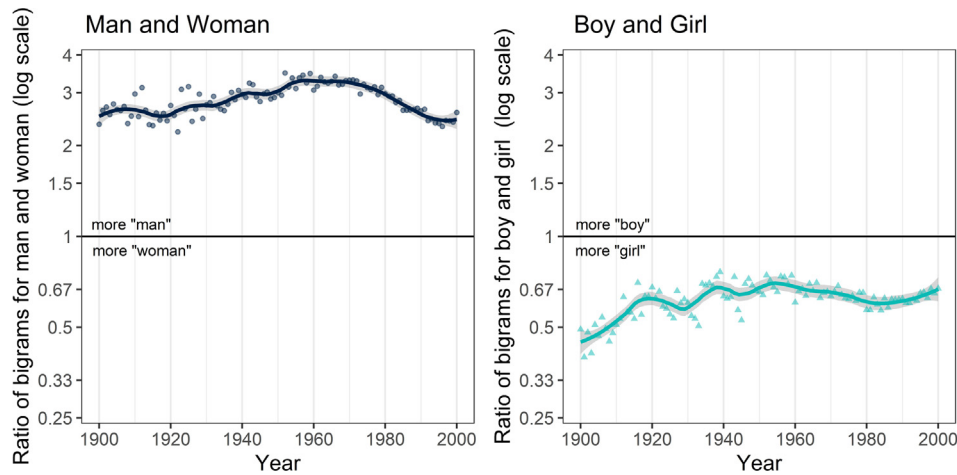
3. Results

Throughout the studied period of time there were more bigrams describing men than women and more bigrams describing girls than boys (see Fig. 1).

To model the trends in the description of genders in the literature, we used a segmented regression for each noun (or difference between two nouns) and dimension (Muggeo, 2008). The number of breakpoints in each model was selected using the Bayesian information criterion (BIC). Afterward, a segmented regression model was fitted with the number of breakpoints determined by the BIC. The resulting models for the difference between the genders are displayed in Fig. 2 and models for the two genders separately are displayed in Fig. 3.

Both genders were generally described in positive terms; that is, well above the midpoint of the scale (3). Men ($M = 4.09$, $SD = 0.04$) were generally depicted in more positive terms than women ($M = 3.99$, $SD = 0.04$), $t(198) = -19.09$, $p < .001$, $b = -0.10$, 95% CI = [−0.11, −0.09]. The positivity of adjectives used to describe men was also less variable ($M = 1.14$, $SD = 0.02$) than the positivity of adjectives used to describe women ($M = 1.24$, $SD = 0.03$), $t(198) = 29.94$, $p < .001$, $b = 0.106$, 95% CI = [0.099, 0.113] (Fig. 4). Men were more likely to be depicted as "honest" and less likely to be described as "foolish", "unhappy", "jealous", "vulgar", and "silly", which most influenced the higher average positivity of description of men. On the other hand, men were more likely to be described as "lazy" and "mean" and less likely to be described as "charming", which had the opposite, but overall weaker, influence (see Table 2). The difference in positivity decreased until about 1937 and increased back afterward. While men were depicted in similarly positive terms throughout most of the 20th century (1913–1989) with an exception of the decrease at the beginning and increase at the end of the century, positivity of description of women was decreasing throughout the century.

Both genders were described in slightly feminine terms; that is, on average below the midpoint of the scale (3). There was generally little difference in masculinity of adjectives used for description of men ($M = 2.86$, $SD = 0.01$) and women ($M = 2.86$, $SD = 0.02$), $t(198) = -0.64$, $p = .53$, $b = -0.00$, 95% CI = [−0.01, 0.00], with a possible exception of the end of the century when men tended to be described in somewhat less masculine adjectives than women. However, masculinity of adjectives used to describe men was less variable ($M = 0.61$, $SD = 0.03$) than masculinity of adjectives used to describe women ($M = 0.63$, $SD = 0.03$), $t(198) = 10.15$, $p < .001$, $b = 0.023$, 95% CI = [0.019, 0.028]. The variability also tended to increase throughout most of the century for both genders (Fig. 4). Men were more likely to be described as "wise", "honorable", and "able", all relatively masculine words, and less likely to be described as "fashionable" and "warm", both relatively

Fig. 1. Ratio of bigrams describing men and women (left) and boys and girls (right). The lines show results of loess regression fitted for the relative frequencies of bigrams. The shaded regions represent 95% confidence intervals for the regression line. Note that the ordinate uses a logarithmic scale.
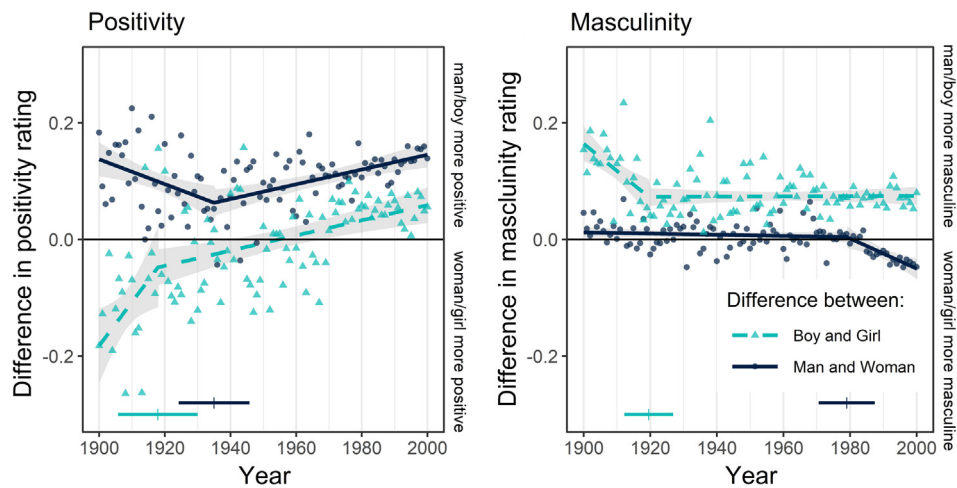
Fig. 2. The difference in positivity and masculinity associated with adjectives used for description of men and women, and of boys and girls. The lines represent results of segmented regressions and shaded regions show 95% confidence intervals for the regression lines. Lines at the bottom of the graph show 95% confidence intervals for the breakpoints.

feminine words. On the other hand, women were more likely to be described as "charming", "vulgar", "independent", and "foolish", all rated as relatively masculine traits. Men were described in decreasingly masculine terms throughout the beginning (1900–1937) and end (1975–2000) of the 20th century with little change between these two periods. Masculinity of adjectives used to describe women did not change throughout the first two-thirds of the century (1900–1968), but it started to slightly increase afterward (1968–2000).

Positivity of adjectives showed the same general trends for both boys and girls. While they were described in increasingly more positive terms until around 1938, the positivity was slightly decreasing from then on. The increase in positivity at the beginning of the century was larger for boys and the subsequent decrease was smaller for boys, resulting in a reversal of the relative positivity of adjectives used in relation to boys and girls. While girls were described more positively at the beginning of the century, the difference was changing in favor of boys throughout the studied period and boys started to be described in more positive terms since around 1960. The relative change of the difference was especially prominent at the beginning of the 20th century (1900–1918). In aggregate, boys (M = 4.04, SD = 0.12) were

described in similarly positive terms as girls (M = 4.06, SD = 0.07), $t(198)$ = 1.41, p = .16, b = 0.02, 95% CI = [−0.01, 0.04]. However, the positivity of adjectives used to describe boys was more variable (M = 1.24, SD = 0.08) than positivity of adjectives used to describe girls (M = 1.15, SD = 0.05), $t(198)$ = −11.17, p < .001, b = −0.092, 95% CI = [−0.108, −0.075]. The variability decreased for both boys and girls until about 1938 and was increasing back afterward. Boys were more likely to be described as "good", which had the most influence on the difference in positivity of descriptions of boys and girls given the overall frequency of the corresponding bigrams (31.7% for boys and 22.5% for girls). To a smaller degree, boys were depicted more positively also because they were more likely to be described as "bright" and less likely to be described as "unhappy" and "inexperienced". On the other hand, boys were more likely to be described by the negative adjectives "foolish", "dull", and "cruel" and less likely to be described by the positive adjectives "nice", "charming", and "sensible" which decreased the difference in positivity of description of the two genders (Table 2).

Children were depicted in more feminine terms than adults. Boys (M = 2.73, SD = 0.04) were described generally in more masculine adjectives than girls (M = 2.65, SD = 0.03), $t(198)$ = −19.25, p < .001, b = −0.08, 95% CI = [−0.09, −0.07]. The variability of
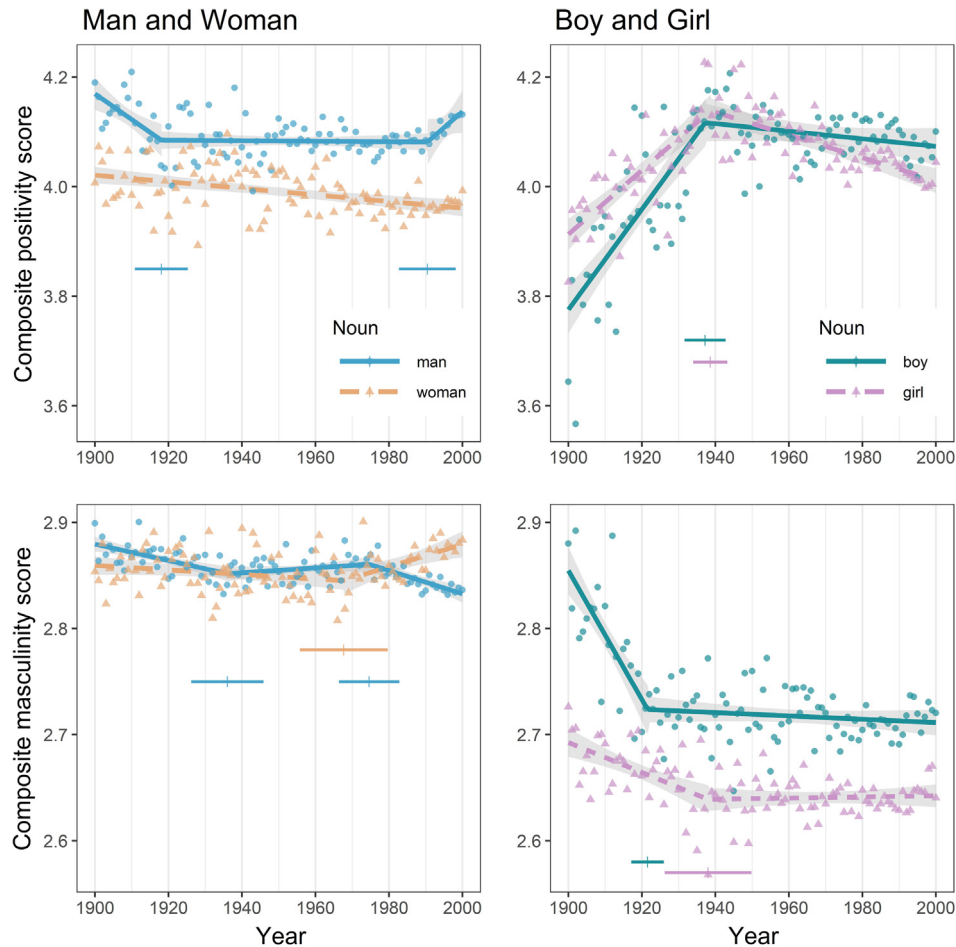
Fig. 3. Positivity and masculinity associated with adjectives used for description of men, women, boys, and girls. The lines represent results of segmented regressions and shaded regions show 95% confidence intervals for the regression lines. Horizontal lines below the data points show 95% confidence intervals for the breakpoints. Note that the ordinate differs between the graphs for positivity and masculinity.
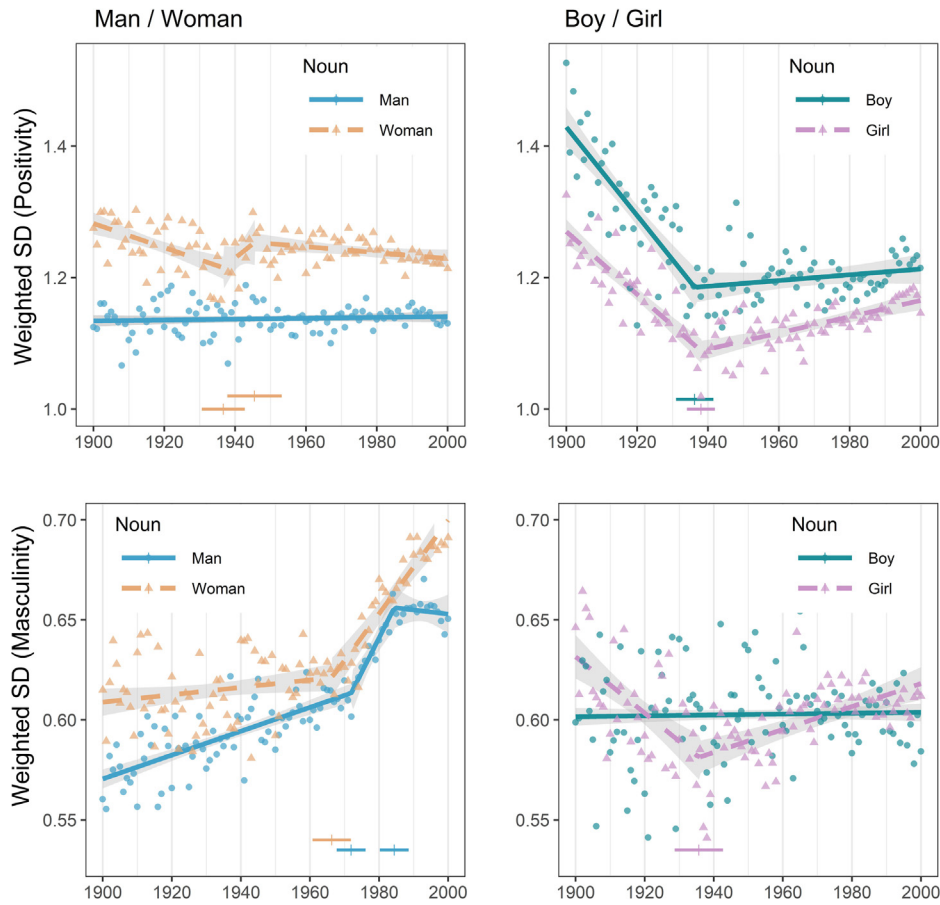
masculinity of adjectives used to describe boys ($M = 0.60$, $SD = 0.02$) and girls ($M = 0.60$, $SD = 0.02$) did not differ appreciably, $t(198) = -0.11$, $p = .91$, $b = -0.000$, 95% CI = [-0.006, 0.006]. There was no general trend for the variability of masculinity of adjectives used to describe boys, but variability of adjectives used to describe girls was decreasing until 1936 and increased afterward. Boys were more likely to be depicted as "tough", "bright", "reckless", and "cruel", all rated as relatively masculine traits, and less likely to be described as "innocent" and "nice", both rated as feminine. On the other hand, girls were more likely to be described as "charming", rated as relatively masculine, and less likely to be described as "good" and "sensitive", both rated as feminine traits. The difference in masculinity of descriptions decreased at the beginning of the 20th century (1900–1920) and then stayed at the same level over the studied period of time. Masculinity of adjectives used for description of boys and girls displayed a similar development in time. While it was decreasing until about 1922 for boys and 1938 for girls, it did not change appreciably until the end of the century.

Finally, we examined the similarity of adjectives used to describe characters of the two genders. Namely, we computed Euclidean distance of the proportions of adjectives used to describe pairs of nouns (see Fig. 5). The analysis showed that men and women ($M = 0.07$, $SD = 0.01$) were generally described in more similar terms than boys and girls ($M = 0.11$, $SD = 0.02$), $t(198) = 16.60$, $p < .001$, $b = 0.040$, 95% CI = [0.035, 0.045]. While the similarity of description of girls and boys did not change appreciably during the century, men and women were described increasingly

more similarly until about 1977 when the trend reversed. Women ($M = 0.17$, $SD = 0.05$) were described generally more similarly to "person" than men ($M = 0.18$, $SD = 0.04$), $t(198) = 2.56$, $p = .01$, $b = 0.005$, 95% CI = [0.001, 0.009], but the difference actually occurred only in the second half of the century. Moreover, the terms used to describe "child" were more similar to the terms used to describe girls ($M = 0.20$, $SD = 0.03$) than to the terms used to describe boys ($M = 0.25$, $SD = 0.04$), $t(198) = 14.50$, $p < .001$, $b = 0.053$, 95% CI = [0.046, 0.060]. Similarly, the terms used to describe "person" were more similar to the terms used to describe girls ($M = 0.23$, $SD = 0.03$) than to the terms used to describe boys ($M = 0.29$, $SD = 0.03$), $t(198) = 18.10$, $p < .001$, $b = 0.067$, 95% CI = [0.060, 0.074]. While boys and girls tended to be described increasingly less similarly as "child" toward the end of the twentieth century, the similarity to description of "person" actually increased. In fact, at the end of the century, both girls and boys were described in terms more similar to "person" than to "child". Finally, women and girls ($M = 0.14$, $SD = 0.02$) tended to be described more similarly than men and boys ($M = 0.18$, $SD = 0.02$), $t(198) = 16.25$, $p < .001$, $b = 0.039$, 95% CI = [0.034, 0.044], but the difference seemed to disappear at the end of the century.

4. Discussion

Similarly as previous studies examining children's books and other media, we found that men were more often portrayed in the English-language fiction than women. As McCabe et al.

Fig. 4. Variability in positivity and masculinity of adjectives used for description of men, women, boys, and girls. The lines represent results of segmented regressions and shaded regions show 95% confidence intervals for the regression lines. Horizontal lines below the data points show 95% confidence intervals for the breakpoints. Note that the ordinate differs between the graphs for positivity and masculinity.

Table 2
The adjectives that had most influence on the overall differences between the two genders. The table displays ten adjectives with highest influence scores for all combinations of the two pairs of nouns and the two dimensions. The influence score shows how much would the difference in average composite scores for a given dimension (shown in the last row) change if a given adjective was not included in computation of the composite scores. The letters in brackets show which of the two nouns was relatively more often associated with the adjective.

Rank	Positivity		Masculinity	
	Man-woman	Boy-girl	Man-woman	Boy-girl
1st	Honest [M] (−0.0246)	Good [B] (−0.1475)	Charming [W] (+0.017)	Good [B] (+0.0278)
2nd	Foolish [W] (−0.0239)	Nice [G] (+0.0199)	Wise [M] (−0.0100)	Charming [G] (+0.0262)
3rd	Unhappy [W] (−0.0174)	Bright [B] (−0.0199)	Vulgar [W] (+0.0091)	Innocent [G] (−0.0195)
4th	Charming [W] (+0.0168)	Charming [G] (+0.0163)	Innocent [M] (+0.0087)	Nice [G] (−0.0101)
5th	Jealous [W] (−0.0148)	Foolish [B] (+0.0161)	Fashionable [W] (−0.0077)	Sensitive [B] (+0.0089)
6th	Vulgar [W] (−0.0131)	Unhappy [G] (−0.0155)	Independent [W] (+0.0073)	Tough [B] (−0.0056)
7th	Good [M] (+0.0127)	Dull [B] (+0.0124)	Honorable [M] (−0.0057)	Bright [B] (−0.0048)
8th	Silly [W] (−0.0110)	Sensible [G] (+0.0103)	Able [M] (−0.0055)	Proud [G] (+0.0047)
9th	Lazy [M] (+0.0099)	Inexperienced [G] (−0.0102)	Warm [W] (−0.0052)	Reckless [B] (−0.0038)
10th	Mean [M] (+0.0098)	Cruel [B] (+0.0100)	Foolish [W] (+0.0046)	Cruel [B] (−0.0035)
Absolute difference	0.1032	−0.0165	0.0014	0.0831

(2011) observed for children's books and Twenge et al. (2012b) observed for the relative use of gendered pronouns in U.S. books, the difference was lower at the beginning and end of the 20th cen-

tury, but unlike in children's books the frequency of portrayal of men in general fiction was always at least twice as high as that of women. The same pattern of results for characters depicted in
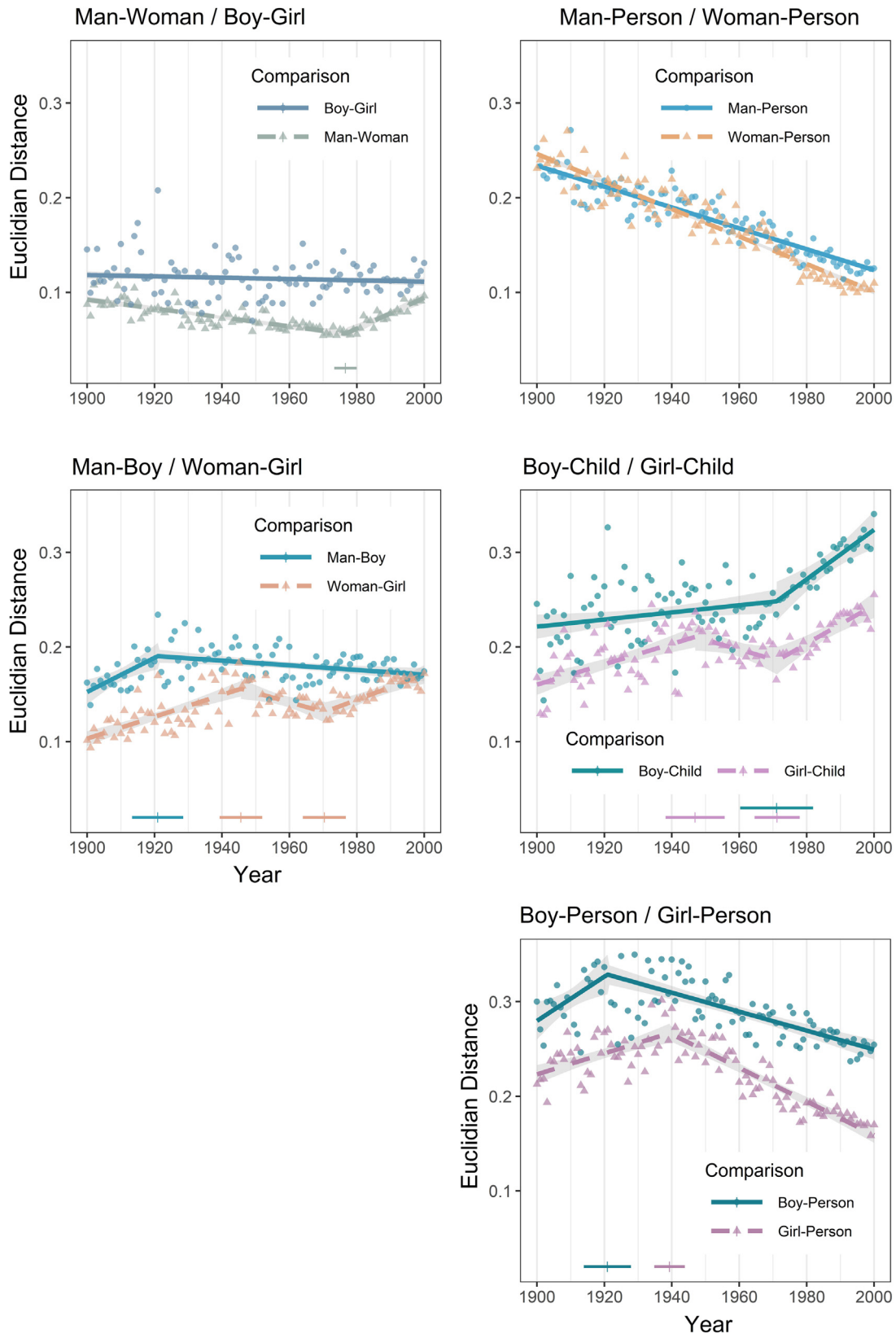
Fig. 5. Euclidean distances of proportions of adjectives used for description of pairs of nouns. The lines represent results of segmented regressions and shaded regions show 95% confidence intervals for the regression lines. Horizontal lines below the data points show 95% confidence intervals for the breakpoints.

books has been recently identified by Underwood et al. (2018) using a different method for identifying characters and a different source of data, supporting validity of the approach used in the present study. Underwood et al. (2018) argue that the observed pattern may be caused by the proportion of women authors which was at its lowest around 1970. Women tend to include women characters relatively more often, so the nadir of women authors around 1970 could explain the higher male-to-female ratio of

characters in books. Here, we show that the overrepresentation of male characters did not hold for children, where "boy" was less likely to occur in a bigram than "girl". It is possible that children characters are more likely to have only a supporting role and that the tendency of male authors to write predominantly about male characters is limited to main roles and is thus less likely to influence the relative frequencies of children characters of the two genders. Yet, given that the adult nouns were much more frequent than the children nouns, male characters were represented in the books more often than female characters overall. Future studies may explore whether the difference between adult and children characters is present also in other media.

While some previous research found that women are evaluated more positively than men (Eagly et al., 1991), we observed the opposite result in portrayal of men and women in the literature. Men were generally described in more positive terms. The difference also did not decrease throughout the century. Men were still described more positively than women even at the end of the century despite the, in many aspects, increased status of women in the society. As a speculation, it is possible that male authors tend to describe men relatively more positively and the higher number of male authors could thus cause the relatively more positive description of male characters. It is also possible that male characters are more likely to be given central roles and main characters are more likely to be described positively than supporting characters. The pattern of results differed for girls and boys. While girls were described in more positive terms at the beginning of the century, the difference gradually reversed throughout the century.

The lack of a difference in the average masculinity of adjectives describing men and women seems surprising. One possible reason for this finding may be that feminine characteristics of men and masculine characteristics of women may be more likely to be mentioned because they are more likely to be informative. Describing a feature of a person that is expected is not as informative as describing a feature that goes against expectations and which therefore distinguishes the person better. For example, "male nurse" is used more often in English-language books than "female nurse" even though nurses are more likely to be female than male. However, the expected difference in masculinity was observed for boys and girls, which suggests that the association of the adjectives with masculinity tracks a real feature. It is also noteworthy that adjectives used for description of children were generally less associated with masculinity than adjectives used for description of adults. It has been found that children are rated to have less agency (Gray, Gray, & Wegner, 2007), which could have been tracked by the masculinity dimension (Spence & Helmreich, 1980).

The interpretation of trends in the depiction of the two genders is not only complicated by the effect of informativeness of adjectives, which may influence the relative frequency of their use for describing the two genders, but also by the difference in meaning of adjectives when they are used to describe men and women. For example, the height necessary for people to be referred to as "tall man" and "tall woman" might be quite different. It is similarly possible that "angry woman" and "angry man" refer to different levels of anger. This could explain the difference between self-reported traits in actual people and traits by which men and women are described in the literature (Ye et al., 2018). Insofar that the dimension of masculinity-femininity tracks how much a given adjective is stereotypically viewed as associated with men or women, the lack of a difference between depiction of men and women on the masculinity dimension would suggest that depiction of the two genders in the literature does not reflect contemporary stereotypes very well, at least for adult characters. However, it is still possible that, while the differences between the genders are hard to interpret, trends in adjectives used in association with a single noun might be meaningful. For example, women were described more

often as "independent", a relatively masculine adjective, than men. The difference in the use of the adjective would therefore not correspond with the stereotype that men are more independent than women. Yet, looking at the trend of use of the bigram "independent woman" shows that women started to be described much more as "independent" since around 1970, when they also gained more economical as well as social independence. The trend of the use of the bigram in the literature could thus be meaningfully related to a real-world change even if the difference between the use of "independent" in association with the two genders is not.

Underwood et al. (2018) showed that gender of fiction characters became harder to classify based on their description and associated words over the 20th century. This could suggest that characters in books became less stereotypically associated with their own genders. We did not see a similar convergence in masculinity of adjectives used to depict men and women in our data because there was no difference in their masculinity from the beginning of the 20th century. Only in children nouns did we observe a closing gap between genders, but it could be seen only at the beginning of the 20th century until about 1920. However, the analysis of similarity of description of men and women showed increasing similarity for most of the 20th century, which could partly explain the increasing difficulty of classification of characters based on gender observed by Underwood et al. The increased similarity of description of men and women might correspond to increased participation of women in the workforce and consequent decreased difference in social roles of men and women. While the stable similarity of the description of boys and girls would be in accord with this explanation, it would not readily explain the decreasing similarity of portrayal of men and women in the last quarter of the century.

Even though it is possible to study the use of words and combinations of words in the literature using the Google Books Ngram corpus, it is often difficult to interpret what the results mean. For example, the observed differences in frequencies of bigrams using "man" and "woman" or "boy" and "girl" suggest overrepresentation of male adult characters and female children characters in the literature; however, it is not clear whether the difference is caused by the number of characters of both genders, their prominency in books, or both. Examining only the development of frequencies of bigrams without their historical context also means that it is hard to interpret what was the cause of changes in representation of the genders. The main results are therefore mostly descriptions of trends and their explanations have to be studied differently.

The lack of specific tested effects of historical events means that we examined only long-term trends rather than short-term fluctuations. It is possible that some short periods of time (e.g., the world wars) significantly influenced the depiction of characters in the literature, but the effect was obscured by the analysis we used. It is therefore not necessarily true that the results we describe apply in the same manner to the whole studied period. On the other hand, it is possible that most short-term events are not likely to influence the depiction of characters immediately, because unlike other media books take longer to appear from the start of their inception, and the short-term events would therefore not show a discernible effect anyway. Bentley, Acerbi, Ormerod, and Lampos (2014), for example, argue that the condition of economy influences language used in the literature most with a ten-year delay.

The composite scores of positivity and masculinity were computed from ratings of adjectives from the perspective of people living in the 21st century. The comparison of ratings of desirability from almost half a century ago (Anderson, 1968) with ratings obtained in the present study showed a near-perfect correlation, suggesting that the word associations are largely stable; yet, insofar that the meaning of some of the words changed (Pettit, 2016),

the composite scores for farther past could be biased. Using a list of words from Anderson (1968) also means that some of the deprecated older words as well as some novel words could not have been included even if they were frequently used in a certain time period.

The corpus does not contain all published fiction books and it is not known to what degree are the included books representative of all the books (Koplenig, 2017). Furthermore, the fiction corpus does not contain only works of fiction, but also some works associated with fiction such as commentaries (http://www.culturomics.org/Resources/faq#dataquality). However, a small proportion of texts that do not fall in the fiction category is unlikely to largely influence our results. The fiction corpus has been shown to be less influenced by publishing trends than the general corpus, which is significantly influenced by changing proportion of scientific texts (Pechenick, Danforth, & Dodds, 2015). While the changing proportion of fiction books on all books may not be an issue for the present study, the changing composition of different genres of fiction may be behind some of the results. Some genres may be more likely to include male or female characters and portray them differently than other genres. The observed trends might then be caused by the changing composition of the corpus in terms of literary genres rather than by a change of portrayal of characters within any given genre. Furthermore, the results do not necessarily correspond to the change of depiction of characters that readers encountered during a given time period. Each book is in the corpus only once, independent of its popularity (Pechenick et al., 2015). More popular books that had larger influence on gender perception could have differed in their portrayal of characters from less popular books which had lesser influence on the culture at a given time period. The characters that the readers actually read about would then differ from the average characters depicted in all books.

Even though the interpretation of the results is associated with certain limitations, the present research shows a possible approach for studying the depiction of the two genders in the literature. Unlike most of the previous studies using the Google Books Ngram corpus, our study did not pick specific n-grams for analysis in an ad-hoc fashion, and it is therefore not influenced by bias in selection of particular words or phrases as some previous studies could have been. While previous research studying depiction of men and women in various media reported results mostly consistent with stereotypical views of the two genders, we did not find evidence that such stereotypes are reflected in adjectives used to describe characters of the two genders in fiction books. In particular, we found that men were described in relatively more positive terms than women, but that the terms did not differ in their association with masculinity. These results suggest that there is no direct correspondence of stereotypes and depiction of characters in books as could have been predicted based on the results of existing studies. The analysis of similarity of adjectives used to describe the selected nouns showed that men and women were described increasingly more similarly until about 1977 when the trend reversed. In the second half of the century, children were described increasingly more similarly to "person" and less similarly to "child". Future studies could look into these and other reported results and search for their explanations. The limitations of the present study also mean that the results should be replicated using different data or methods that could corroborate the current results.

Acknowledgements

Author contributions

Both DS and ŠB designed the study and wrote the report. DS analyzed the data.

References

Acerbi, A., Lampos, V., Garnett, P., & Bentley, R. A. (2013). The expression of emotions in 20th century books. *PLOS ONE, 8*(3), e59030.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*(1), i–171.

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9*(3), 272–279.

Bentley, R. A., Acerbi, A., Ormerod, P., & Lampos, V. (2014). Books average previous decade of economic misery. *PLOS ONE, 9*(1), e83147.

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin, 26*(10), 1171–1188.

Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? *Psychology of Women Quarterly, 15*(2), 203–216.

Eisend, M. (2010). A meta-analysis of gender roles in advertising. *Journal of the Academy of Marketing Science, 38*(4), 418–440.

Elasmar, M., Hasegawa, K., & Brain, M. (1999). The portrayal of women in US prime time television. *Journal of Broadcasting & Electronic Media, 43*(1), 20–34.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83.

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues, 55*(3), 473–489.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology, 60*, 115–140.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619.

Kite, M. E., Deaux, K., & Haines, E. L. (2008). Gender stereotypes. In F. L. Denmark & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (2nd ed., pp. 205–236). Westport, CT: Praeger.

Koplenig, A. (2017). The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets—Reconstructing the composition of the German corpus in times of WWII. *Digital Scholarship in the Humanities, 32*(1), 169–188.

Lauzen, M. M., & Dozier, D. M. (2005). Maintaining the double standard: Portrayals of age and gender in popular films. *Sex Roles, 52*(7), 437–446.

Lewis, M., & Lupyan, G. (2019). Language use shapes cultural stereotypes: Large scale evidence from gender. https://doi.org/10.31234/osf.io/7qd3g.

Lin, Y., Michel, J. B., Aiden, E. L., Orwant, J., Brockman, W., & Petrov, S. (2012). Syntactic annotations for the Google books ngram corpus. In *Proceedings of the ACL 2012 system demonstrations* (pp. 169–174). Association for Computational Linguistics.

Lupyan, G., & Lewis, M. (2017). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience, 1–19*. https://doi.org/10.1080/23273798.2017.1404114.

Mason, S. E., Kuntz, C. V., & McGill, C. M. (2015). Oldsters and Ngrams: Age stereotypes across time. *Psychological Reports, 116*(1), 324–329.

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*(2), 197–226.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176–182.

Muggeo, V. M. R. (2008). segmented: An R package to fit regression models with broken-line relationships. *R News, 8*(1), 20–25.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology, 60*, 339–367.

Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLOS ONE, 10*(10), e0137041.

Pettit, M. (2016). Historical time in the age of big data: Cultural psychology, historical change, and the Google Books Ngram Viewer. *History of Psychology, 19*(2), 141–153.

Spence, J. T., & Helmreich, R. L. (1980). Masculine instrumentality and feminine expressiveness: Their relationships with sex role attitudes and behaviors. *Psychology of Women Quarterly, 5*(2), 147–163.

Suitner, C., & Maass, A. (2008). The role of valence in the perception of agency and communion. *European Journal of Social Psychology, 38*(7), 1073–1082.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66*(1), 21–36.

Thompson, T. L., & Zerbinos, E. (1995). Gender roles in animated cartoons: Has the picture changed in 20 years? *Sex Roles, 32*(9–10), 651–673.

Turner-Bowker, D. M. (1996). Gender stereotyped descriptors in children's picture books: Does "curious Jane" exist in the literature? *Sex Roles, 35*(7), 461–488.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012b). Male and female pronoun use in US books reflects women's status, 1900–2008. *Sex Roles, 67*(9–10), 488–493.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012a). Increases in individualistic words and phrases in American books, 1960–2008. *PLOS ONE, 7*(7), e40181.

Underwood, T., Bamman, D., & Lee, S. (2018). The transformation of gender in English-language fiction. *Journal of Cultural Analytics.* https://doi.org/10.22148/16.019.

Ye, S., Cai, S., Chen, C., Wan, Q., & Qian, X. (2018). How have males and females been described over the past two centuries? An analysis of Big-Five personality-related adjectives in the Google English Books. *Journal of Research in Personality, 76*, 6–16.