Check for updates

Research Article

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
1–15
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09567976251349815
www.psychologicalscience.org/PS

Sage

Syntactic and Semantic Gender Biases in the Language on Children's Television: Evidence From a Corpus of 98 Shows From 1960 to 2018

Andrea C. Vial[1], Aida Mostafazadeh Davani[2], Ruyuan Zuo[2], Shreya Havaldar[2], Eleanor K. Chestnut[3], Morteza Dehghani[2,4], and Andrei Cimpian[3]

[1]Psychology Program, Science Division, New York University Abu Dhabi; [2]Department of Computer Science, University of Southern California; [3]Department of Psychology, New York University; and [4]Department of Psychology, University of Southern California

Abstract

Biased media content shapes children's social concepts and identities. We examined gender bias in a large corpus of scripts from 98 children's television programs from the United States spanning the years 1960 to 2018 (6,600 episodes, ~2.7 million sentences, ~16 million words). We focused on agency and communion, the fundamental psychological dimensions underlying gender stereotypes. At the syntactic level, words referring to men or boys (vs. women or girls) appear more often in the agent (vs. patient) role. This syntactic bias remained stable between 1960 and 2018. At the semantic level, words referring to men or boys (vs. women or girls) co-occurred more often with words denoting agency. Words denoting communion showed both stereotypical and counterstereotypical associations. Some semantic gender biases have remained unchanged or have weakened over time; others have grown. These findings suggest that gender stereotypes are built into the core of children's stories. Whether we are closer today to gender equality in children's media depends on where one looks.

Keywords

agency, communion, gender, media, natural language processing

Received 8/23/24; Revision accepted 5/21/25

Stereotyped media content shapes children's beliefs about gender (e.g., what women and men are like) and their self-views, interests, and motivations (for a review, see Ward & Grower, 2020). Cultural stereotypes associate men and boys with *agency* (e.g., being competent, assertive, and in control) and portray them as low on *communion* (e.g., having aggressive or violent tendencies; Eagly et al., 2020). In contrast, women and girls are viewed as highly communal (e.g., caring, nurturing) but low in agency (e.g., needy; Wakefield et al., 2012). These views are transmitted to children through various sources, and mass media—television in particular—have long been identified as a crucial conduit for stereotypes (Mutz & Goldman, 2010). We used natural language processing (NLP) techniques to examine whether the fundamental psychological dimensions of agency and communion (Fiske et al., 2002) are portrayed in gender-stereotyped ways in the language of a large corpus of scripts from children's television programs that aired in the United States over the last six decades, including not only its semantic content but also its syntactic structure.

Whether through broadcast TV, cable TV, or streaming, children and adolescents consume vast amounts of television (Common Sense Media, 2020, 2021). Media

Corresponding Author:
Andrea C. Vial, Psychology Program, Science Division, New York University Abu Dhabi
Email: andrea.vial@nyu.edu

products for children are cultural artifacts reflecting and promoting a given sociocultural view (Dehghani et al., 2013; Morling & Lamoreaux, 2008). They constitute a key part of a system of social representations (Moscovici, 1988) that connect (macro) historical and cultural processes and (micro) psychological phenomena, such as gender stereotyping (Lamer et al., 2022; Lloyd & Duveen, 1990). Television's power to cultivate viewers' beliefs (Gerbner et al., 2002) is complemented by its ability to supply models for young viewers to emulate (Mutz & Goldman, 2010). If television programming for children consistently portrays female characters as primarily communal and male characters as primarily agentic, then children who watch these programs will be more likely to see such patterns as characteristic of women and men in general. Consequently, girls may veer away from activities that emphasize agency and toward ones that emphasize communion, whereas boys may do the opposite (Evans & Diekman, 2009). Thus, quantifying the presence of gender stereotypes in children's media is highly consequential, given substantial evidence of its power to shape children's attitudes (Ward & Grower, 2020).

Traditionally, gender stereotypes in children's media have been examined via content-analytic methods that focus on visual elements and plotlines from limited samples (Collins, 2011; Diekman & Murnen, 2004; Ward & Grower, 2020). Large-scale analysis of the language in these programs, however, remains notably scarce. Linguistic cues such as word co-occurrence and sentence structure reliably convey socially relevant information (Kinzler, 2021), including gender stereotypes and norms (e.g., Chestnut & Markman, 2018), and because of their subtlety, they are less likely to be recognized or challenged than stereotypes encoded on a surface level into the programs (e.g., the actions or demographics of the characters). The rising popularity of script-writing programs powered by artificial intelligence (AI), which are trained on language from pre-existing screenplays (Dayo et al., 2023), adds urgency to the goal of uncovering social biases in the language in children's media. We focused precisely on these below-the-surface linguistic cues, taking advantage of NLP methods to examine a large corpus of children's media.

In recent years, NLP techniques have been applied to other corpora. These studies, however, are limited, either because they did not focus on media *designed for children* (Caliskan et al., 2022; Formanowicz et al., 2017; Garg et al., 2018), which is arguably most relevant to understanding the intergenerational transmission of stereotypes, or because they focused exclusively on semantic content (Charlesworth et al., 2021; Gálvez et al., 2019; Lewis et al., 2022). These studies overlooked the fact that *syntax* also powerfully conveys gender stereotypes (Chestnut & Markman, 2018). For example, the syntactic structure of a question such as, "Do women lead differently than men?" communicates that men (not women) are the cultural standard in the leadership domain (Formanowicz & Hansen, 2022). Syntax shapes social beliefs (e.g., Kesebir, 2017), including information about agency (who are the "doers"?) and passivity (who are the "done-tos"?; Kako, 2006; Maass et al., 2014). However, gender stereotypes at the syntactic level in large corpora of text aimed at children remain unexplored. We fill this gap by analyzing for the first time, to our knowledge, gender biases not only in the semantic content of children's media but also in its syntactic structure.

We used NLP tools to compare how often words referring to men or boys and women or girls[1] appeared as the syntactic agents versus the patients in a sentence. A syntactic bias to place men in agent roles and women in patient roles has been found in smaller, specialized text corpora such as linguistics textbooks (Kotek et al., 2021), but to our knowledge it has yet to be examined in cultural artifacts with broad reach. We tested this *syntactic-bias hypothesis*, expecting to find that male words would take on the agent role more frequently than female words, mirroring the stereotype that agency is primarily the province of men (Eagly et al., 2020; Wakefield et al., 2012). We complemented this content-free syntactic analysis with an analysis of stereotypes in semantic content. Specifically, we analyzed the co-occurrence of male and female words with words related to agency (e.g., "power," "money") and communion (e.g., "family," "friend"). A word's linguistic context is informative about its meaning (Boleda, 2020; Firth, 1957), so the words with which gender words tend to co-occur are informative about their semantic content (that is, gender stereotypes). This analysis enabled us to test a *semantic-bias hypothesis*: namely, that male (vs. female) words would co-occur more often with words denoting agency, whereas female (vs. male) words would co-occur more often with words denoting communion. Although syntactic structure and semantic content are not directly comparable, both may convey gender biases that influence the development of gender cognitions in children.

Given the longitudinal nature of our data set, we explored whether these syntactic and semantic biases changed over the last six decades. This period has marked a rapid increase in women's participation in labor-market roles that project agency (e.g., lawyers; Lippa et al., 2014), and some gender stereotypes seem to have changed as a result (Eagly et al., 2020). However, evidence that mainstream media reflect such improvements remains mixed (Gálvez et al., 2019; Jones et al., 2020). It is possible that media products change

mostly on the surface (e.g., converging proportions of female and male characters) without deeper changes in the linguistic patterns that reinforce the ideological foundation of gender dynamics (e.g., men act, women appear; Collins, 2011; Diekman & Murnen, 2004). Moreover, whereas biases in semantic content may be relatively transparent and thus more amenable to conscious revision by screenwriters responding to cultural change, syntactic biases—though equally impactful psychologically (e.g., Chestnut & Markman, 2018; Rhodes et al., 2019)—are woven into the very structure of language, making them more resistant to deliberate intervention.

Research Transparency Statement

General disclosures

Conflicts of interest: All authors declare no conflicts of interest. **Funding:** This research was supported by National Science Foundation Grant No. BCS-1733897 and by Mattel, Inc.'s Dream Gap Postdoctoral Fellowship. **Artificial intelligence:** Artificial intelligence technologies (Claude 3.5) were used to assist with the creation of R code to conduct cross-validation tests. No other artificial-intelligence-assisted technologies were used in this research or in the creation of this article. **Ethics:** This research received approval from the ethics board at New York University (IRB-FY2016-1163). **Open Science Framework (OSF):** To facilitate long-term preservation, we have registered all OSF files at https://osf.io/bca9z/.

Study disclosures

Preregistration: No aspects of the study were preregistered. **Materials:** All study materials are publicly available (https://osf.io/bca9z/). **Data:** All primary data are publicly available (https://osf.io/bca9z/). **Analysis scripts:** All analysis scripts are publicly available (https://osf.io/bca9z/). **Computational reproducibility:** The computational reproducibility of the results in the main article (but not the Supplemental Material) has been independently verified by the Journal's STAR team.

Method

We gathered a data set of scripts from children's television programs, which we coded using NLP tools for (a) the frequency of gender words, (b) the syntactic roles of these words, and (c) their co-occurrence with words denoting agency and communion. Agency is often conceptualized as encompassing assertiveness and competence as two distinct but related components (Abele & Wojciszke, 2014). For the syntactic analysis we focused solely on assertiveness (as the agent–patient syntactic role distinction is unrelated to competence), but for the semantic analysis we focused on categories of words tapping both assertiveness and competence. We coded the date of release of each episode to examine change over time, and we coded the genre and target age of the audience to examine the robustness of our results across program characteristics. Moreover, we examined whether our conclusions were robust across three different NLP approaches. To identify syntactic roles, we initially relied on spaCy (Version 3.3.1), an open-source, widely used NLP library in Python (Honnibal & Montani, 2018). We then created two alternative versions of our data set using NLP tools with underlying architectures that differ from spaCy. First, we used the label attention layer (LAL) model, which has a neural network architecture that improves accuracy in analyzing sentence structure; it incorporates specialized attention mechanisms that help identify grammatical relationships between words (Mrini et al., 2020). Using only syntactic tags on which the two models (spaCy and LAL) agreed strengthens confidence in the accuracy of our parsing. Second, we used the power-connotations frame model introduced by Sap et al. (2017), which helps reveal implicit power relationships that simple syntactic tagging might miss (e.g., agent roles that lack agency, patient roles that have agency). This model complements spaCy's automated tagging by providing a richer contextual analysis of agency beyond strict syntactic roles.

Content selection

We scraped the full contents of a large online database (www.SpringfieldSpringfield.co.uk), which included 4,244 programs from the United States for adult and junior audiences. We narrowed down this list to child-oriented programming by conducting a search on Google.com using different keywords by decade (e.g., "Kid TV shows 1960s"). Ambiguous cases were resolved by consulting the Internet Movie Database (www.IMDb.com), a large online repository of detailed information on movies and television programs. When available, the official television rating was consulted (e.g., TV-G). For most programs, we had access to scripts for multiple episodes across multiple seasons. In total, we analyzed text from 98 programs, including 6,603 episodes comprising approximately 2.7 million sentences and 16.2 million words. The full list of programs is reported in the Supplemental Material available online (see Table S1), along with the number of episodes per program and release years.

Main variables of interest

We cleaned the text data for each script by removing all unnecessary punctuation marks and non-ASCII characters. Then, we employed the sentence segmentation function from spaCy (Version 3.3.1) to divide the text of each script into a list of sentences. We then coded (a) whether a sentence contained gender words, (b) what syntactic role each gender word occupied (to test for a syntactic bias), and (c) whether a sentence contained words related to agency, communion, or both (to test for a semantic bias).

Word gender. We employed the *female* and *male* word categories in LIWC2015 (Pennebaker et al., 2015), which contain 124 and 116 words, respectively, to identify gender words. These word categories consist of both pronouns (e.g., "she," "her," "he," "his") and nouns (e.g., "woman," "girl," "man," "boy"). We identified 351,224 gender words (123,662 female words and 227,562 male words) contained in 313,700 unique sentences (comprising 2,625,043 words in total) across 6,600 episodes.

Syntactic roles and their links to agency. We used the sentence tokenizer and parser functions from spaCy to identify words in agent versus patient grammatical roles. The agent role in a sentence corresponds to the participant that initiates or sustains an action, whereas the patient role is taken by the participant upon whom actions are carried out. For example, in the sentence "The boys fed the cats," "boys" refers to the agent and "cats" refers to the patient. Conceptually, agent roles and patient roles are inversely related, but this relationship is not perfect (e.g., a sentence could have an agent without a patient, as in "The cats are eating"). We considered words with the "nominal subject" or "agent" dependency labels in spaCy as agents, and words with the "direct object," "indirect object," or "passive nominal subject" dependency labels in spaCy as patients. In cases in which the agent is not in the subject position because of passive construction, spaCy accurately detects the agent as preceded by the preposition "by" (e.g., in the sentence, "The cats were fed by the boys," spaCy would tag "the boys" as the agent). Words that did not meet these criteria were labeled as "other." The spaCy model labeled 116,230 of the gender words we identified in the previous step as "agent," 66,555 gender words as "patient," and 168,439 gender words as "other." (In the section below titled "Alternative coding methods," we describe an additional step we took to verify that syntactic agents and patients did in fact display actions consistent with agent and patient roles, respectively.)

We computed the ratio of per-episode frequencies of agent roles and patient roles separately for female versus male words. We added 1 to the numerator and denominator to prevent ratios from having undefined values.

Semantic content related to agency and communion. We coded the agency- and communion-related semantic content of the television scripts in two different ways: (a) with the 41 word categories denoting psychological processes in the LIWC2015 software and (b) with recently validated word lists that were custom-built to capture agency- and communion-related content (Pietraszkiewicz et al., 2019).

To identify which of the 41 LIWC2015 word categories are relevant to agency and/or communion, we collected data from adults on Amazon Mechanical Turk (MTurk; $N = 47$; 48.9% female, 72.3% White; the study was approved by the Institutional Review Board at New York University). Many of the word categories include lower-order sets of words—for example, the *drives* word category has five lower-order categories, including *affiliation* and *power*. We only employed lower-order word categories to avoid overlap between categories. The full set of 41 word categories, along with word examples for each, is listed in the Supplemental Material, Table S4. We should note that if a female word (e.g., "aunt") appears in a word category (e.g., *family*), that category also includes a corresponding male word (e.g., "uncle"), if available. Thus, the differential patterns of co-occurrence between male and female words and words in the agency- and communion-related word categories is not simply an artifact of the composition of these lists.

In brief, for each word category, MTurk participants saw the category label (e.g., *positive emotion*) followed by words in that category ("funniness," "opportun*," "cheerful," etc.), in random order. Participants then indicated whether the words in the category were related to (in counterbalanced order) (a) agency and, separately, (b) communion on a scale ranging from 0 (*totally unrelated*) to 100 (*closely related*). We defined agency as "an orientation to success and self-expansion, a focus that is separated from others, a striving to act and reach goals, and to be active and efficient," and communion as "an orientation to people, a focus on interpersonal contacts and relationships, a striving for being included into a community, and for remaining a member of this community" (Abele & Wojciszke, 2014). These definitions were always visible (for additional details, see Section B and Fig. S1 in the Supplemental Material).

We considered any category with either agency or communion ratings above the scale midpoint, and for which agency and communion ratings differed significantly, to be exclusively related either to agency or to communion, depending on the ratings. With this strategy, we identified six word categories exclusively

related to agency (*achievement*, *causation*, *health*, *money*, *power*, and *reward*) and four word categories exclusively related to communion (*affiliation*, *family*, *friends*, and *religion*). We considered any category with both agency and communion ratings above the scale midpoints, but whose ratings did not differ significantly in agency versus communion, to be related to both agency and communion. Three word categories met these criteria (*home*, *work*, and *positive emotion*). The 28 remaining word categories did not meet these classification criteria and were not analyzed further. The full results are reported in the Supplemental Material in Section C and Tables S5 and S6.

We coded each sentence in our data set using the LIWC2015 categories described above as well as Pietraszkiewicz et al.'s (2019) validated lists of words with agentic and communal content. For each word list or category, each sentence was assigned a score of 0 (no words from the list/category were present in the sentence) or 1 (some words from the list/category were present in the sentence). Our analysis entailed counting the number of sentences coded as 1 for each word category or list that also included a gender word. For example, the sentence "Never send a boy to do a man's job" (*Bewitched*, Season 1, Episode 10, 1964) was coded as containing a word from the LIWC word category *work* ("job") co-occurring with male words ("boy," "man"). The sentence "She needs to stay in bed for a few days" (*My Little Pony: Friendship is Magic*, Season 2, Episode 16, 2012) was coded as containing a word from the LIWC word category *home* ("bed") co-occurring with a female word ("she"). The proportions of sentences with female and male words for each word list or category are reported in Table S4 in the Supplemental Material. This method provides a straightforward means of examining the semantic content of gender words: A word's meaning can be inferred from its linguistic context (Boleda, 2020; Firth, 1957). Thus, we can use information about which words occur in the same sentence as male and female words to make reliable inferences about the meaning of these words (that is, gender stereotypes).

Episode date of release and tests for first-order temporal autocorrelation. We obtained the air date for each episode, ranging from September 30, 1960, to November 11, 2018, from IMDb.com. We conducted a Durbin-Watson test (King, 1983) to detect potential serial dependence in our data, whereby word frequencies measured closer in time might be more strongly related to each other than word frequencies in distant time points (which can lead to biased or misleading model coefficients). The Durbin-Watson statistic ranges between 0 and 4; the closer a value is to 2, the less evidence there is for first-order autocorrelation. The test, which was performed on the spaCy data set at the episode level (syntactic-bias analysis), revealed a Durbin-Watson value of 2.01, which was not significantly different from 2.00, $p = .656$, indicating no serial dependence. Whereas temporal autocorrelation might be expected in aggregated annual data, its absence here is reasonable because our data are at the episode level (i.e., very granular). Natural episode-to-episode variation in content (e.g., more action-focused vs. dialogue-focused) creates fluctuations in language that can easily overshadow temporal correlations.

Covariates of interest and alternative coding methods

To assess the robustness of our results, we examined whether our conclusions were also supported (a) when considering differences between the programs in terms of genre and target age and (b) when using alternative NLP tools to code the data.

Covariates: program genre and target age. We gathered information on program characteristics from IMDb.com and CommonSenseMedia.org, including whether a program was a family program (87.8%) or not and whether it was animated (42.9%) or not, and we coded up to three genres for each program: comedy (77.8% of programs); action (25.5%); adventure (42.9%); fantasy (18.4%); and drama (33.7%). We searched each program on CommonSenseMedia.org and followed the website's recommendation to record the minimum age for which the program's content was considered developmentally appropriate. In three cases, this information was not available; for two of these three cases, we based our coding on the programs' available rating on IMDb.com. In our sample, 12.2% of programs were for children younger than 6 years of age, 31.6% were for 6- and 7-year-olds, 33.7% were for 8- and 9-year-olds, 11.2% were for 10- and 11-year-olds, and 7.1% were for 12- to 14-year-olds. We coded programs as being for younger children ($n = 52$) if they received a rating of TV-G or TV-Y (i.e., appropriate for all ages or all children), and we coded programs as being for older children ($n = 45$) if they received any other rating (e.g., TV-Y7, TV-14, TV-PG). Program characteristics are reported in full in Table S1 in the Supplemental Material.

Alternative coding methods. We created two additional versions of our data set using alternative NLP methodologies to validate the conclusions drawn from the initial spaCy model. First, we used the LAL model, a state-of-the-art open-source NLP tool that, unlike spaCy's pipeline-based approach, uses neural networks to better

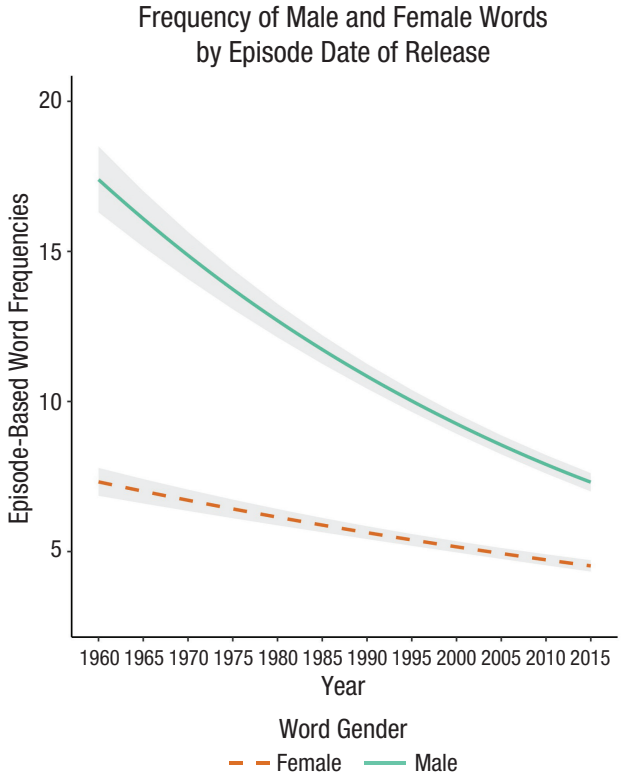

Fig. 1. Per-episode frequency of female words (dashed orange line) and male words (solid green line) in children's television programs as a function of year of release. Error bands represent ±1 *SE*.

handle complex or ambiguous grammatical relationships between words (Mrini et al., 2020). We identified 258,880 sentences for which the LAL and spaCy models converged in their tagging, resulting in 104,728 gender words coded by both models as "agent," 55,764 gender words coded as "patient," and 118,358 gender words coded as "other."

Second, we created a third version of our data set using an NLP tool that directly measures power- and agency-related connotations of verbs (i.e., whether a verb implies that the agent has or lacks power, or that an agent has power over the patient or vice versa; Sap et al., 2017). For example, in the sentence "She needs help," the verb "needs" implies that the agent ("she") has low agency. Similarly, in the sentence "He heard her," the verb connecting the agent ("he") and the patient ("her") implies that the patient in the sentence has more power than the agent. For each gender word tagged as an agent or patient by the spaCy model, we used this tool to derive its agency and power connotations, which we then used to filter out agent roles that lacked agency and patient roles that had agency. After this filtering step, which was applied to the initial spaCy data set, we were left with 82,863 gender words tagged as "agent," 63,540 gender words tagged as "patient," and 168,439 gender words tagged as "other."

Results

Frequency of male and female words

We first tested whether male words were more frequent overall than female words in our corpus. We conducted a mixed-effects Poisson regression, using the R package *lme4* (Version 1.1-35.2; Bates et al., 2015), on the per-episode frequency of sentences that contained a gender word. As fixed effects, the model included word gender (0 = *female* vs. 1 = *male*), the release date of each episode (both mean-centered), and their interaction. As random effects, we included random intercepts for program and episode (nested within program). We also included an offset term to adjust for episode length (i.e., the number of sentences). For these analyses, we report the incidence rate ratio (*IRR*), a ratio of the incidence of one event type (e.g., sentences that contain male words) to the incidence of another event type (e.g., sentences that contain female words).

Results revealed that male words (M = 9.48 words per episode, SE = 0.34) were 1.81 times as common as female words (M = 5.23, SE = 0.19), b = 0.60, SE = 0.004, $p < .001$, IRR = 1.81, 95% confidence interval (CI) = [1.802, 1.827]. This means that, in a typical episode in our data set, a child watching would hear roughly double the number of male words than female words. The model also revealed a significant interaction between word gender and episode date of release, b = −0.12, SE = 0.003, $p < .001$, IRR = 0.89, 95% CI = [0.885, 0.897], indicating that the gender gap decreased about 11% each year. As seen in Figure 1, an analysis of the simple slopes revealed that the frequency of gender words declined overall, but male words declined by about 23% each year, b = −0.26, SE = 0.02, $p < .001$, IRR = 0.77, 95% CI = [0.740, 0.806], whereas female words declined less sharply (i.e., by about 13% each year), b = −0.14, SE = 0.02, $p < .001$, IRR = 0.87, 95% CI = [0.830, 0.904]. A child watching in the late 2010s would still hear about 50% more male words than female words in a typical episode.

Differential occurrence of male and female words in agent and patient syntactic roles

To test the syntactic-bias hypothesis—namely, that male (vs. female) words would more often be the syntactic agents (vs. patients)—we submitted the per-episode word frequencies to a mixed-effects Poisson regression with the following fixed effects: word gender (0 = female words vs. 1 = male words), syntactic role (0 = patient roles vs. 1 = agent roles), the release date of each episode (all mean-centered), and all possible

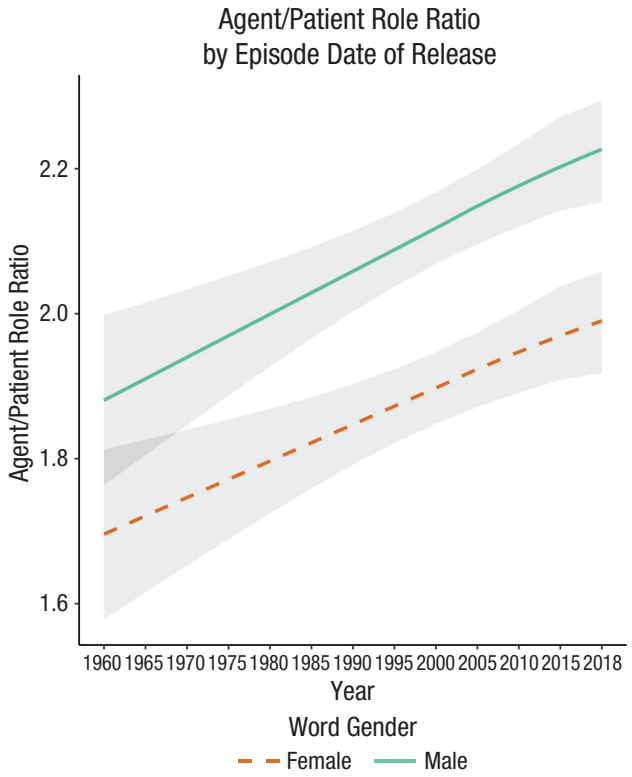

Fig. 2. Per-episode agent:patient ratio for female words (dashed orange line) and male words (solid green line) across time. Error bands represent ±1 *SE*.

interactions. As random effects, we included random intercepts for program and episode (nested within program). An offset term for episode length (i.e., number of sentences) was also included. Gender words (female or male) appeared significantly more often overall in agent roles ($M = 6.73$, $SE = 0.29$) than in patient roles ($M = 3.84$, $SE = 0.17$), $b = 0.56$, $SE = 0.01$, $p < .001$, $IRR = 1.75$, 95% CI = [1.735, 1.770]. This difference was moderated by word gender, $b = 0.04$, $SE = 0.01$, $p < .001$, $IRR = 1.04$, 95% CI = [1.017, 1.058]. Consistent with the syntactic-bias hypothesis, the difference favoring agent (vs. patient) roles was larger for male words, $b = 0.58$, $SE = 0.01$, $p < .001$, $IRR = 1.79$, 95% CI = [1.764, 1.807], than for female words, $b = 0.54$, $SE = 0.01$, $p < .001$, $IRR = 1.72$, 95% CI = [1.694, 1.747]. That is, although gender words were generally more likely to be found in the agent (vs. patient) role, this was particularly the case for male words, as expected. Looking at this interaction another way, male words appeared more often than female words in both syntactic roles, but this gender gap was larger for agent roles, $b = 0.55$, $SE = 0.01$, $p < .001$, $IRR = 1.74$, 95% CI = [1.720, 1.761], than for patient roles, $b = 0.52$, $SE = 0.01$, $p < .001$, $IRR = 1.68$, 95% CI = [1.651, 1.704]. For every 10 times that someone was described as a "doer," this was a male character about 6.4 times and a female character about 3.6 times. For "done-tos," the gender gap was in the same direction but slightly smaller (6.3 vs. 3.7 times, respectively).

The three-way interaction between word gender, syntactic role, and episode date of release was not significant, $b = 0.01$, $SE = 0.01$, $p = .375$, $IRR = 1.01$, 95% CI = [0.989, 1.029]. Figure 2 plots the ratio of per-episode frequencies of agent roles versus patient roles separately for female and male words. As seen in the figure, the agent:patient ratio was significantly larger for male words ($M = 2.11$, $SE = 0.05$) than for female words ($M = 1.89$, $SE = 0.05$), $b = 0.22$, 95% CI = [0.167, 0.271], $SE = 0.03$, $p < .001$, and this gap remained stable. Thus, even though male words declined in frequency over time (Fig. 1), they did so in a way that simultaneously maintained the conflation of agency with males at the syntactic level.

Differential co-occurrence of male and female words with words denoting agency and communion

Next, we tested the semantic-bias hypothesis—namely, that male and female words would be disproportionately associated with semantic content denoting agency and communion, respectively. We focused on the 13 LIWC2015 word categories that, on the basis of our pretest, we classified as exclusively related to agency (*achievement, cause, health, money, power*, and *reward*), exclusively related to communion (*affiliation, family, friend, religion*), or related to both (*home*, *work*, *positive emotion*). We assessed the co-occurrence (within a sentence) of female versus male words with words contained in each of these 13 categories.

Given that the probability that words from any particular word category would appear in any particular sentence was low (i.e., the 0s greatly outnumbered the 1s in our dependent variables), which can bias estimation of parameters in a logistic regression (Leevy et al., 2018), we created 100 random subsets of the data for each word category with equal numbers of 0s (i.e., no words in the category included in a sentence) and 1s (i.e., some words in the category included in a sentence), and we conducted 100 iterations of a mixed-effects logistic regression for each word category. The dependent variable in each model was whether or not a sentence contained a word from the relevant category. As fixed effects, each model included word gender (0 = the sentence contained a female word vs. 1 = the sentence contained a male word), the episode's date of release (both mean-centered), and their interaction, as well as the syntactic role of the gender word as a

covariate (which adjusted for the previously described syntactic gender bias). As random effects, each model included random intercepts for program and episode (nested within program). An offset term for sentence length was also included.

The only word categories retained for further consideration were those for which at least 95% of the models on balanced subsets revealed a significant difference by word gender or a significant interaction between word gender and episode date of release (see Section A and Table S2 in the Supplemental Material). This included three agency-related categories (*power*, *money*, and *reward*), three communion-related categories (*affiliation*, *family*, and *friend*), and two categories related to both agency and communion (*home* and *work*). For these eight word categories, we conducted the same mixed-effects logistic regression on the entire imbalanced data set. To prevent multiple comparisons from inflating the Type I error rate, we used the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) with a 5% false discovery rate to identify which differences remained significant when accounting for the number of models (see Section A and Table S3 in the Supplemental Material).

In addition to these LIWC2015 categories, we used two previously validated word lists (Pietraszkiewicz et al., 2019) that capture *agency* and *communion* globally. We followed a similar analytic strategy as with the LIWC2015 word categories. Below, we report coefficients, *p* values, odds ratios (*OR*s), and confidence intervals from these models, grouped by dimension (agency, communion, both). We start with the two global word lists (Pietraszkiewicz et al., 2019) before presenting the results for more specific LIWC2015 categories related to these dimensions. Moreover, as a robustness check, given that 24 female words that were identified in our data do not have an equivalent word in the LIWC2015 male dictionary (and vice versa for 9 male words), which could potentially bias the results, we repeated the analysis on a subset of our data set that excluded any sentences containing gender words without an equivalent. Specifically, 8,615 sentences were excluded, leaving 342,609 sentences across 6,600 episodes from 98 programs. The full results of these ancillary tests are reported in the Supplemental Material (Section G, Table S21, and Fig. S4). With one exception, the results converged with those based on the full data set for all word categories and lists; we note the exception below.

Agency-related words. Male words (vs. female words) co-occurred more often with words in the global *agency* word list (Pietraszkiewicz et al., 2019), $b = 0.05$, $SE = 0.01$, $p < .001$, $OR = 1.05$, 95% CI = [1.031, 1.064]. When agency-related words appeared in a sentence, the odds of finding a male word in that same sentence were 5% higher than the odds of finding a female word. As seen in Figure 3a, this gender gap in words related to *agency* grew over time, $b = 0.03$, $SE = 0.01$, $p < .001$, $OR = 1.04$, 95% CI = [1.020, 1.052]. Specifically, the gap in the odds of finding a male versus female word in the same sentence as an agency-related word increased by about 4% each year over the span of our data. This increase in the agency gender gap was observed because associations between agency and male words increased over time, $b = 0.04$, $SE = 0.02$, $p = .008$, $OR = 1.04$, 95% CI = [1.011, 1.077], whereas associations between agency and female words did not, $b = 0.01$, $SE = 0.02$, $p = .653$, $OR = 1.01$, 95% CI = [0.975, 1.042].

We now turn to the narrower agency-related word categories from LIWC2015. Male words (vs. female words) co-occurred more often with words related to *money* (e.g., "bucks," "income"), $b = 0.21$, $SE = 0.02$, $p < .001$, $OR = 1.24$, 95% CI = [1.190, 1.289], and with words related to *reward* (e.g., "win," "success"), $b = 0.07$, $SE = 0.01$, $p < .001$, $OR = 1.07$, 95% CI = [1.049, 1.095]. When money-related words appeared in a sentence, the odds of finding a male word in that same sentence were 24% higher than the odds of finding a female word; similarly, for reward-related words, the odds were 7% higher for male words than female words. These gaps did not change over time. Male words (vs. female words) also co-occurred more often with words related to *power* (e.g., "strong," "respected"), $b = 0.34$, $SE = 0.01$, $p < .001$, $OR = 1.40$, 95% CI = [1.376, 1.430]. When power-related words appeared in a sentence, the odds of finding a male word in that same sentence were 40% higher than the odds of finding a female word. However, as seen in Figure 3b, the gender gap in words related to *power* decreased over time, $b = -0.10$, $SE = 0.01$, $p < .001$, $OR = 0.90$, 95% CI = [0.886, 0.919]. Specifically, the gap in the odds of finding a male versus female word in the same sentence as a power-related word decreased by 10% each year. This decrease was due to a significant decrease in associations between power and male words, $b = -0.08$, $SE = 0.03$, $p = .014$, $OR = 0.93$, 95% CI = [0.869, 0.984]. Associations between power and female words did not change significantly, $b = 0.02$, $SE = 0.03$, $p = .448$, $OR = 1.03$, 95% CI = [0.962, 1.092].

Taken together, these patterns illuminate an overall male bias in agency-related semantic content that may be growing (i.e., given the results on the global *agency* word list), even though some semantic subcategories (e.g., words related to *power*) exhibit decreasing gender gaps.

Communion-related words. Semantic content related to communion revealed somewhat surprising, counter-stereotypical patterns. First, contrary to our expectations,

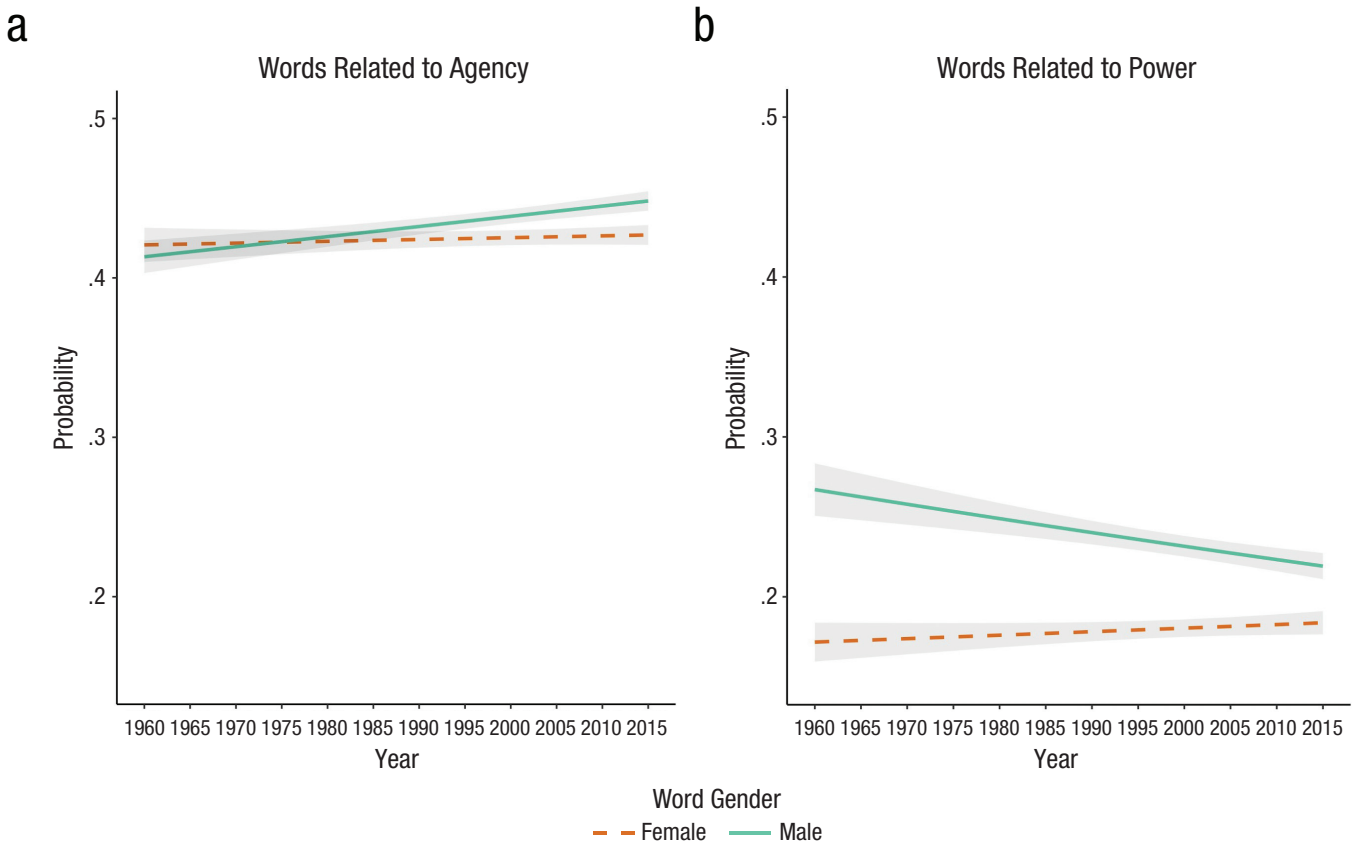

Fig. 3. The estimated probability that sentences with female words (dashed orange line) and male words (solid green line) also contained agency words (a) and power words (b), as a function of time. Error bands represent ±1 *SE*.

male words (vs. female words) co-occurred more often with words in the global *communion* word list (Pietraszkiewicz et al., 2019), $b = 0.07$, $SE = 0.01$, $p < .001$, $OR = 1.07$, 95% CI = [1.049, 1.091]. When communion-related words appeared in a sentence, the odds of finding a male word in that same sentence were 7% higher than the odds of finding a female word. This male bias persisted over time (see Fig. 4a). It should be noted, however, that we found no gender bias on the global *communion* word list in the subset of our data set that excluded gender words without an equivalent in the other LIWC2015 gender dictionary. Thus, the male bias in the global *communion* word list should be interpreted with caution, as it may be a byproduct of subtle imbalances in our female and male word lists.

When looking at specific communion-related word categories from LIWC2015, we found a surprising pattern for content related to *friends* that mirrors the unexpected male bias found with the global *communion* word list. Specifically, words denoting *friends* (e.g., "pal," "roommate") co-occurred significantly more frequently with male than female words, $b = 1.79$, $SE = 0.02$, $p < .001$, $OR = 6.00$, 95% CI = [5.797, 6.216]. This difference has increased over time, $b = 0.51$, $SE = 0.02$, $p < .001$, $OR = 1.66$, 95% CI = [1.603, 1.723], as seen in Figure 4b, because of a simultaneous decrease in co-occurrence with female words, $b = -0.25$, $SE = 0.05$, $p < .001$, $OR = 0.78$, 95% CI = [0.697, 0.863], and a growing association with male words, $b = 0.25$, $SE = 0.05$, $p < .001$, $OR = 1.29$, 95% CI = [1.164, 1.428]. Of note, although a close inspection of the *friends* word category revealed that it contained a number of friendship-related words that seem distinctly male or masculine (e.g., "buddy," "chum"), with no female or feminine counterparts, the conclusions did not change when we excluded gender words without equivalents (see the Supplemental Material, Section G). This friends–male association could reflect a high representation of programs (or episodes within programs) featuring many male characters (Bechdel, 1986), particularly in more recent years.

Beyond these counterstereotypical associations, we found the expected female biases in other word categories related to communion, but these biases appear to have declined. Female words (vs. male words) co-occurred more often with words related to *affiliation* (e.g., "relationship," "loyal"), $b = -0.04$, $SE = 0.01$, $p < .001$, $OR = 0.96$, 95% CI = [0.944, 0.981], and *family*

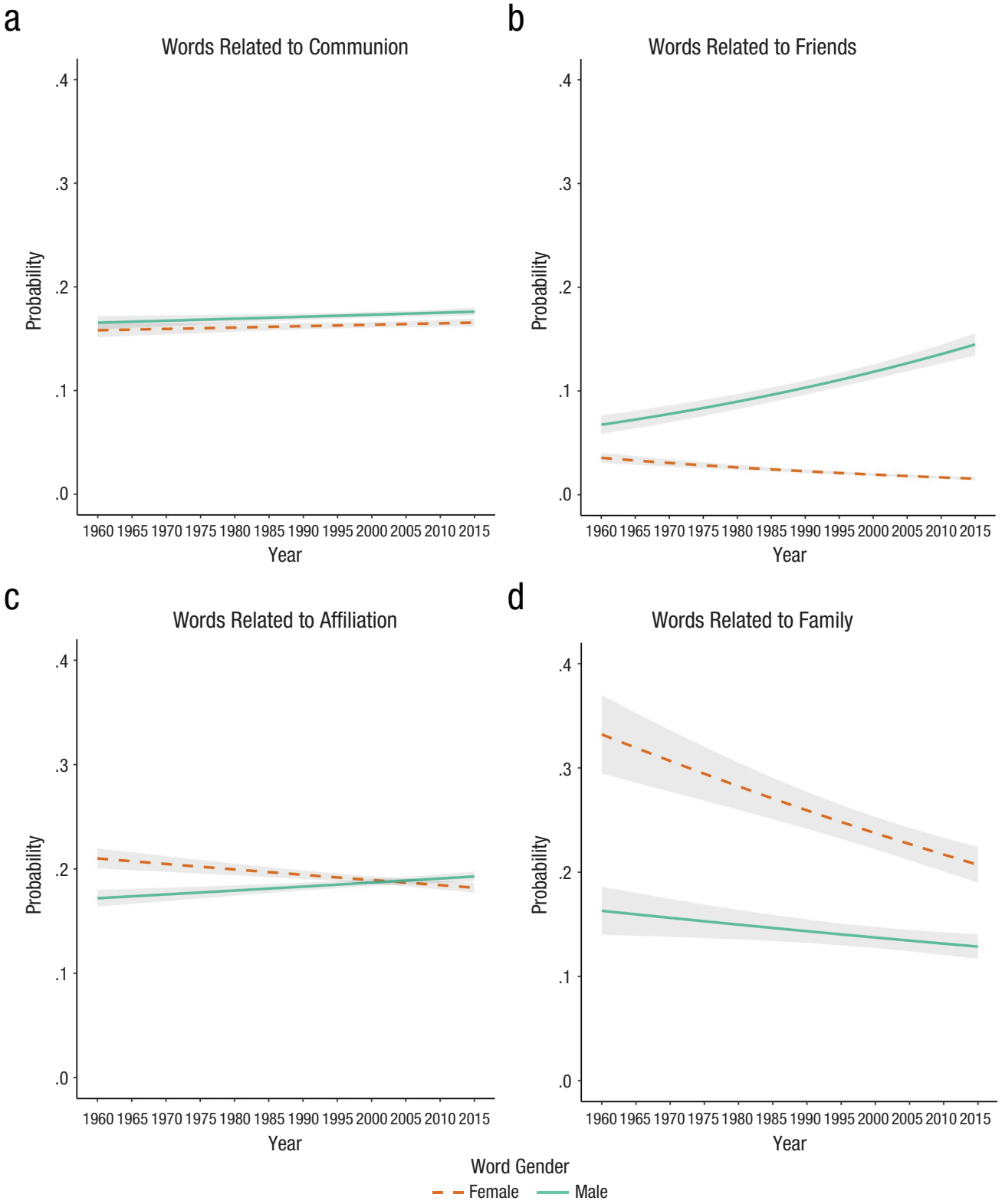

Fig. 4. The estimated probability that sentences with female words (dashed orange line) and male words (solid green line) also contained words related to communion (a), friends (b), affiliation (c), and family (d) as a function of time. Error bands represent ±1 *SE*.

(e.g., "wedding," "son"), $b = -0.70$, $SE = 0.01$, $p < .001$, $OR = 0.50$, 95% CI = [0.489, 0.508]. The odds of finding a male word in a sentence were only 96% the odds of finding a female word when affiliation-related words also appeared in that sentence, and the odds of finding a male word in a sentence were 50% when family-related words appeared in that sentence. As illustrated in Figures 4c and 4d, these gender gaps have decreased over time (affiliation: $b = 0.09$, $SE = 0.01$, $p < .001$, $OR = 1.10$, 95% CI = [1.079, 1.121]; family: $b = 0.11$, $SE = 0.01$, $p < .001$, $OR = 1.12$, 95% CI = [1.094, 1.138]). In both cases, decreases in gender gaps were due to significant decreases in associations with female words (affiliation: $b = -0.05$, $SE = 0.02$, $p = .018$, $OR = 0.95$, 95% CI = [0.907, 0.991]; family: $b = -0.19$, $SE = 0.06$, $p = .002$, $OR = 0.83$, 95% CI = [0.731, 0.931]) juxtaposed with stable associations with male words (affiliation: $b = 0.04$, $SE = 0.02$, $p = .054$, $OR = 1.04$, 95% CI = [0.999, 1.088]; family: $b = -0.08$, $SE = 0.06$, $p = .179$, $OR = 0.92$, 95% CI = [0.816, 1.038]).

In sum, across communion-related word categories, we observed a trend toward less gender-stereotypical content, as some of these word categories changed toward less biased representations (*affiliation*, *family*), whereas others showed consistent counterstereotypical associations (*communion*), some of which appear to be growing (*friends*).

Words related to both agency and communion. Two word categories from the LIWC2015 set (*home* and *work*) were rated as being relevant to both agency and communion. However, words related to *home* (e.g., "mop," "porch") co-occurred more often with female than male words, $b = -0.26$, $SE = 0.02$, $p < .001$, $OR = 0.77$, 95% CI = [0.746, 0.803]. Conversely, words related to *work* (e.g., "job," "project") co-occurred more often with male than female words, $b = 0.15$, $SE = 0.01$, $p < .001$, $OR = 1.16$, 95% CI = [1.126, 1.188]. The odds of finding a male word in a sentence were 77% the odds of finding a female word when home-related words also appeared in that sentence, and the odds were 116% when work-related words appeared in that sentence. These gender gaps (Figs. 5a and 5b, respectively) did not change significantly over time. Thus, semantic content related to *home* and *work*, which epitomize traditional gender roles (Eagly & Steffen, 1984), showed stable gender-stereotypic associations.

Robustness check 1: cross-validation analysis

We applied a cross-validation framework to examine the robustness of our results by evaluating model performance on different train-test splits of our data set. We implemented a *k*-fold cross-validation system by creating five folds of the data, each containing 20% of the programs in our data set. The details of this approach, and the full results, are reported in the Supplemental Material (Section H, Tables S22–26). On the whole, the cross-validation results supported our conclusions and indicated that our findings are robust. The only exceptions were as follows. First, although the male bias in overall word frequency was robust, the interaction with episode date of release, which indicated a closing gender gap, was not. Thus, the temporal trend suggesting a growing representation of female (vs. male) words over time should be interpreted with caution. Second, the models suggesting gender bias in use of friendship and family words showed evidence of overfitting, which means that our findings of a closing female bias in content related to family and of a growing (counterstereotypic and unexpected) male bias in content related to friendship may not generalize beyond our data set and should be viewed with caution.

Robustness check 2: gender bias as a function of television-program characteristics

We repeated the analyses above including the following covariates in our models: target audience age, status as a family program, whether the program is animated, and program genre (comedy, action, adventure, fantasy, or drama). The results did not change when these covariates were accounted for (see the Supplemental Material, Section D and Tables S7–S16). Thus, our conclusions appear robust to variations in program characteristics, suggesting that they apply on average across the 98 programs in our data set.

Robustness check 3: alternative coding methods

As described in the Method section, we created two additional versions of our data set using alternative NLP methods to identify the syntactic role of gender words. The full results for these alternative analyses are reported in the Supplemental Material, Sections E (Tables S17 and S18), F (Tables S19 and S20), and G (Table S21). None of our conclusions changed when we employed these alternative NLP methods.

Discussion

The images that children see on television are understood, and thereby shape children's beliefs, via the language that accompanies them. Using NLP techniques,

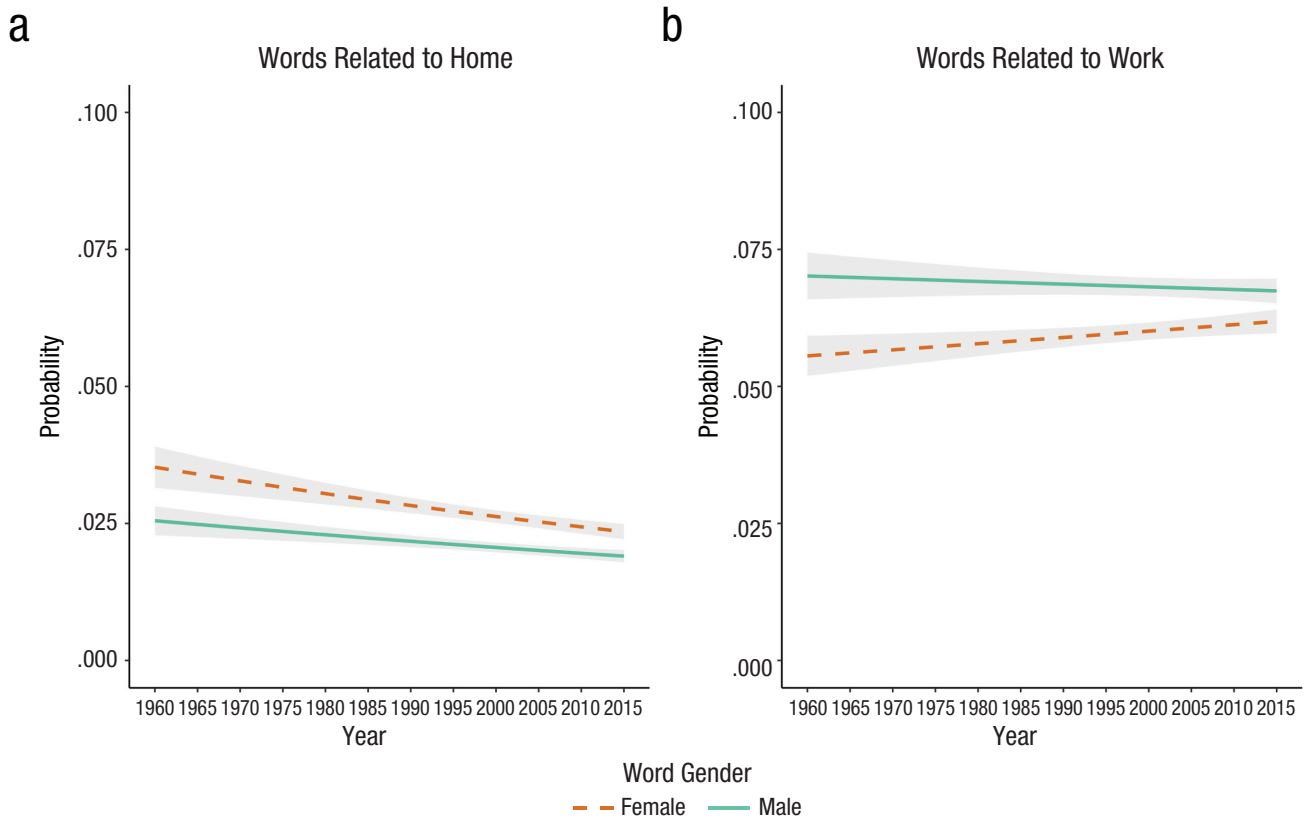

Fig. 5. The estimated probability that sentences with female words (dashed orange line) and male words (solid green line) also contained home words (a) and work words (b), as a function of time. Error bands represent ±1 *SE*.

we analyzed the language on children's television programs to examine whether, and how, it communicates gender stereotypes about agency and communion (Eagly et al., 2020; Wakefield et al., 2012). To our knowledge, the current investigation is the first to employ NLP techniques to examine these dimensions not only in the semantic content of a vast corpus of children's media, but also in its syntactic structure. Our findings illuminate how gender stereotypes are built into the very core of children's stories—including their grammar—and how the extent of progress in how gender is portrayed depends on where one looks.

As in previous investigations using other methods (Charlesworth et al., 2021; Gálvez et al., 2019), we found that female words were less frequent than male words, and that male words were more likely than female words to be associated with semantic content denoting agency. Female words were more likely to be associated with communion—with notable exceptions, such as words related to friendship. (We urge caution when interpreting this finding because of questionable model reliability in our cross-validation analysis.) Although these gaps were sometimes small, their cumulative impact over years of watching likely contributes to the cultivation of a gender-stereotypic world view (Gerbner et al., 2002). Our distinctive contribution, however, lies in uncovering more subtle biases in how men and women are talked about—namely, a *syntactic* gender bias in which male words in our corpus were more likely than female words to be used as the grammatical agent (vs. the patient) in a sentence. This syntactic gender bias communicates in an indirect but pervasive manner traditional gender ideologies about men being the active "doers" in society (Kako, 2006; Maass et al., 2014). Substantial prior evidence indicates that subtle syntactic cues such as these shape children's beliefs about gender (e.g., Chestnut & Markman, 2018; Rhodes et al., 2019). Thus, this small but persistent syntactic gender bias is likely to contribute to the intergenerational transmission of gender stereotypes.

Our analysis of change over time revealed mixed evidence of improvement, with persistent gender biases being found at the syntactic level of analysis. In purely numerical terms, the gap in female representation decreased between 1960 and 2018. Although this development should be viewed with caution given that the corresponding coefficient was somewhat unstable in our cross-validation analyses (see the Supplemental

Material, Section H), it parallels changes documented in other media (Gálvez et al., 2019). At the semantic level, we found mixed patterns of stability and change. Whereas some semantic content showed a (modest) tendency toward less traditional associations over time (e.g., weaker male–power associations), we also found signs of persistent stereotypic content—for example, in female–home and male–work associations (Charlesworth et al., 2021; Eagly & Steffen, 1984). Some of these mixed patterns may reflect model overfitting or sparseness in our co-occurrence data. On the whole, however, the semantic content of children's media seems to mirror the broader evolution of gender attitudes across the last six decades. The results at the syntactic level—where biases often escape notice by creators and audiences alike—were more concerning. There, gender stereotypes about agency showed no sign of improvement: The greater probability of male (vs. female) words occupying agent (vs. patient) roles has persisted unchanged since the 1960s—a pattern that, to our knowledge, we are the first to discover. These results highlight how deep-seated gender ideologies can survive despite good intentions (e.g., efforts to increase the number of female characters). Notably, opaque syntactic biases such as those documented here are also likely to be perpetuated by the rise of AI programs used for screenwriting (Dayo et al., 2023), because these programs are typically trained on existing scripts. Our findings thus underscore the importance of analyzing language at multiple levels—from character counts to syntactic patterns—to fully understand how gender biases are encoded in and transmitted through children's media.

As in past NLP investigations (e.g., Gálvez et al., 2019), we examined linguistic content in media without contextualizing this content on the basis of the gender of the characters that produced it. This is a limitation, because character gender could enrich the meaning of linguistic content and its stereotypic relevance. Similarly, coding first-person statements (which might not have a gender pronoun) may be particularly informative when combined with character gender (e.g., "I'm too emotional to go to work today," as spoken by a woman vs. a man). Future investigations may address these limitations, seeking to combine contextual richness with the large-scale NLP approach. Moreover, although our findings were based on a large and diverse corpus of child-directed television programming and are thus likely to generalize to children's programs beyond our sample, the focus on a single medium (television) is a limitation. It would be fruitful to investigate whether the syntactic and semantic gender biases we uncovered might generalize to other types of media for children (e.g., movies, books). Our findings are also most relevant to U.S. media and may not fully extend to children's media from other cultures or in languages other than English (Dehghani et al., 2013; Morling & Lamoreaux, 2008). Finally, although our conclusions were robust to variations across programs in the age range of the children they were intended for, it may be worthwhile to analyze a larger corpus of media directed toward early childhood, given how early gender stereotyping has been documented (Bian et al., 2017).

Gender representations in media influence the development of children's gender norms and beliefs (Ward & Grower, 2020). NLP techniques can be a powerful tool for analyzing the content of these social representations, from more superficial elements such as simple frequencies down to the basic building blocks of syntax. Our findings reveal that gender stereotypes of agency and communion permeate all of these layers and continue to be woven into the fabric of children's television programs.

Transparency

Action Editor: D. A. Briley
Editor: Simine Vazire

Author Contributions

Andrea C. Vial: Conceptualization; Formal analysis; Investigation; Methodology; Project administration; Visualization; Writing – original draft; Writing – review & editing.
Aida Mostafazadeh Davani: Conceptualization; Formal analysis; Investigation; Methodology; Visualization; Writing – review & editing.
Ruyuan Zuo: Investigation; Writing – review & editing.
Shreya Havaldar: Investigation; Writing – review & editing.
Eleanor K. Chestnut: Conceptualization; Writing – review & editing.
Morteza Dehghani: Conceptualization; Writing – review & editing.
Andrei Cimpian: Conceptualization; Formal analysis; Funding acquisition; Writing – original draft; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was supported by National Science Foundation grant BCS-1733897 and by Mattel, Inc.'s Dream Gap Postdoctoral Fellowship.

Artificial Intelligence

Artificial intelligence technologies (Claude 3.5) were used to assist with the creation of R code to conduct cross-validation tests. No other artificial-intelligence-assisted technologies were used in this research or the creation of this article.

Ethics

This research received approval from the ethics board at New York University (IRB-FY2016-1163).

Open Practices

Preregistration: No aspects of the study were preregistered. Materials: All study materials are publicly available and registered at https://osf.io/bca9z/. Data: All primary data are publicly available (https://osf.io/bca9z/). Analysis scripts: All analysis scripts are publicly available (https://osf.io/bca9z/). Computational reproducibility: The computational reproducibility of the results in the main article (but not the Supplemental Material) has been independently verified by the Journal's STAR team.

ORCID iDs

Andrea C. Vial https://orcid.org/0000-0001-9367-2081
Ruyuan Zuo https://orcid.org/0000-0002-6122-5348
Morteza Dehghani https://orcid.org/0000-0002-9478-4365
Andrei Cimpian https://orcid.org/0000-0002-3553-6097

Acknowledgments

We thank Jun Yen Leung, J. P. Entrocassi, Eun-Sung Chang, Ana Varvara Triff, Mollyrose Napolitano, Angelica Vasquez, Ingrid Friedman, Claire O'Leary, Alessia Seroff, and Nejat Mussa for their assistance with this research. We also thank the members of the Cognitive Development Lab at New York University and the Social Roles and Beliefs Lab at New York University Abu Dhabi for their helpful feedback.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976251349815

Note

1. For ease of exposition, we use the shorthand terms "female words" and "male words" from this point on.

References

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 50, pp. 195–255). Academic Press. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bechdel, A. (1986). *Dykes to watch out for*. Firebrand Books.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391. https://doi.org/10.1126/science.aah6524

Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, *6*(1), 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303

Caliskan, A., Ajay, P. P., Charlesworth, T., Wolf, R., & Banaji, M. R. (2022). *Gender bias in word embeddings: A comprehensive analysis of frequency, syntax, and semantics* [Conference session]. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 156–170). https://doi.org/10.1145/3514094.3534162

Charlesworth, T. E., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychological Science*, *32*(2), 218–240. https://doi.org/10.1177/0956797620963619

Chestnut, E. K., & Markman, E. M. (2018). "Girls are as good as boys at math" implies that boys are probably better: A study of expressions of gender equality. *Cognitive Science*, *42*(6), 2229–2249. https://doi.org/10.1111/cogs.12637

Collins, R. L. (2011). Content analysis of gender roles in media: Where are we now and where should we go? *Sex Roles*, *64*(3–4), 290–298. https://doi.org/10.1007/s11199-010-9929-5

Common Sense Media. (2020). *The Common Sense Census: Media use by kids age zero to eight*. https://www.commonsensemedia.org/sites/default/files/research/report/2020_zero_to_eight_census_final_web.pdf

Common Sense Media. (2021). *The Common Sense Census: Media use by tweens and teens*. https://www.commonsensemedia.org/sites/default/files/research/report/8-18-census-integrated-report-final-web_0.pdf

Dayo, F., Memon, A., & Dharejo, N. (2023). Scriptwriting in the age of AI: Revolutionizing storytelling with artificial intelligence. *Journal of Media & Communication*, *4*(1), 24–38.

Dehghani, M., Bang, M., Medin, D., Marin, A., Leddon, E., & Waxman, S. (2013). Epistemologies in the text of children's books: Native-and non-Native-authored books. *International Journal of Science Education*, *35*(13), 2133–2151. https://doi.org/10.1080/09500693.2013.823675

Diekman, A. B., & Murnen, S. K. (2004). Learning to be little women and little men: The inequitable gender equality of nonsexist children's literature. *Sex Roles*, *50*, 373–385. https://doi.org/10.1023/B:SERS.0000018892.26527.ea

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*(4), 735–754. https://doi.org/10.1037/0022-3514.46.4.735

Evans, A. D., & Diekman, A. B. (2009). On motivated role selection: Gender beliefs, distant goals, and career interest. *Psychology of Women Quarterly*, *33*(3), 235–249. https://doi.org/10.1111/j.1471-6402.2009.01493.x

Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In J. R. Firth (Ed.), *Studies in linguistic analysis* (Repr. 1962 ed.) (pp. 1–32). Blackwell.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Formanowicz, M., & Hansen, K. (2022). Subtle linguistic cues affecting gender in(equality). *Journal of Language and Social Psychology*, *41*(1), 127–147. https://doi.org/10.1177/0261927X211035170

Formanowicz, M., Roessel, J., Suitner, C., & Maass, A. (2017). Verbs as linguistic markers of agency: The social side of grammar. *European Journal of Social Psychology*, *47*(5), 566–579. https://doi.org/10.1002/ejsp.2231

Gálvez, R. H., Tiffenberg, V., & Altszyler, E. (2019). Half a century of stereotyping associations between gender and intellectual ability in films. *Sex Roles*, *81*(9–10), 643–654. https://doi.org/10.1007/s11199-019-01019-x

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Gerbner, G., Gross, L., Morgan, M., Signorielli, N., & Shanahan, J. (2002). Growing up with television: Cultivation processes. In J. Bryant & M. B. Oliver (Eds.), *Media effects: Advances in theory and research* (pp. 43–67). Routledge.

Honnibal, M., & Montani, I. (2018). *spaCy: Industrial-strength natural language processing*. https://spacy.io/

Jones, J. J., Amin, M. R., Kim, J., & Skiena, S. (2020). Stereotypical gender associations in language have decreased over time. *Sociological Science*, *7*, 1–35. https://doi.org/10.15195/v7.a1

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, *101*(1), 1–42. https://doi.org/10.1016/j.cognition.2005.08.002

Kesebir, S. (2017). Word order denotes relevance differences: The case of conjoined phrases with lexical gender. *Journal of Personality and Social Psychology*, *113*(2), 262–279. https://psycnet.apa.org/doi/10.1037/pspi0000094

King, M. L. (1983). Testing for autocorrelation in linear regression models: A survey. In M. L. King & D. E. A. Giles (Eds.), *Specification analysis in the linear model* (pp. 19–73). Routledge.

Kinzler, K. D. (2021). Language as a social cue. *Annual Review of Psychology*, *72*, 241–264. https://doi.org/10.1146/annurev-psych-010418-103034

Kotek, H., Dockum, R., Babinski, S., & Geissler, C. (2021). Gender bias and stereotypes in linguistic example sentences. *Language*, *97*(4), 653–677.

Lamer, S. A., Dvorak, P., Biddle, A. M., Pauker, K., & Weisbuch, M. (2022). The transmission of gender stereotypes through televised patterns of nonverbal bias. *Journal of Personality and Social Psychology*, *123*(6), 1315–1335. https://doi.org/10.1037/pspi0000390

Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A., & Seliya, N. (2018). A survey on addressing high-class imbalance in big data. *Journal of Big Data*, *5*, Article 42. https://doi.org/10.1186/s40537-018-0151-6

Lewis, M., Cooper Borkenhagen, M., Converse, E., Lupyan, G., & Seidenberg, M. S. (2022). What might books be teaching young children about gender? *Psychological Science*, *33*(1), 33–47. https://doi.org/10.1177/09567976211024643

Lippa, R. A., Preston, K., & Penner, J. (2014). Women's representation in 60 occupations from 1972 to 2010: More women in high-status jobs, few women in things-oriented jobs. *PLOS ONE*, *9*(5), Article e95960. https://doi.org/10.1371/journal.pone.0095960

Lloyd, B., & Duveen, G. (1990). A semiotic analysis of the development of social representations of gender. In G. Duveen & B. Lloyd (Eds.), *Social representations and the development of knowledge* (pp. 27–46). Cambridge University Press.

Maass, A., Arcuri, L., & Suitner, C. (2014). Shaping intergroup relations through language. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 157–176). Oxford University Press.

Morling, B., & Lamoreaux, M. (2008). Measuring culture outside the head: A meta-analysis of individualism—collectivism in cultural products. *Personality and Social Psychology Review*, *12*(3), 199–221. https://doi.org/10.1177/1088868308318260

Moscovici, S. (1988). Notes toward a description of social representations. *European Journal of Social Psychology*, *18*(3), 211–250. https://doi.org/10.1002/ejsp.2420180303

Mrini, K., Dernoncourt, F., Tran, Q., Bui, T., Chang, W., & Nakashole, N. (2020). Rethinking self-attention: Towards interpretability in neural parsing. *arXiv:1911.03875v3*. https://doi.org/10.48550/arXiv.1911.03875

Mutz, D. C., & Goldman, S. K. (2010). Mass media. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 241–258). Sage.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. http://hdl.handle.net/2152/31333

Pietraszkiewicz, A., Formanowicz, M., Gustafsson Sendén, M., Boyd, R. L., Sikström, S., & Sczesny, S. (2019). The big two dictionaries: Capturing agency and communion in natural language. *European Journal of Social Psychology*, *49*(5), 871–887. https://doi.org/10.1002/ejsp.2561

Rhodes, M., Leslie, S. J., Yee, K. M., & Saunders, K. (2019). Subtle linguistic cues increase girls' engagement in science. *Psychological Science*, *30*(4), 455–466. https://doi.org/10.1177/0956797618823670

Sap, M., Prasettio, M. C., Holtzman, A., Rashkin, H., & Choi, Y. (2017). *Connotation frames of power and agency in modern films* [Conference session]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 2329–2334). https://aclanthology.org/D17-1247

Wakefield, J. R., Hopkins, N., & Greenwood, R. M. (2012). Thanks, but no thanks: Women's avoidance of help-seeking in the context of a dependency-related stereotype. *Psychology of Women Quarterly*, *36*(4), 423–431. https://doi.org/10.1177/0361684312457659

Ward, L. M., & Grower, P. (2020). Media and the development of gender role stereotypes. *Annual Review of Developmental Psychology*, *2*(1), 177–199. https://doi.org/10.1146/annurev-devpsych-051120-010630