**Online Supplemental Materials**

**Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis**

**of U.S. Public Opinion Polls from 1946 to 2018**

**Contained within this document:**

Appendix S1: References to Included Opinion Polls

Appendix S2: Analytical Strategy for Addressing Effect Size Dependencies

Appendix S3: Additional Details and Analyses Regarding the Equal Response Category

Table S1:Likability Ratings of the Traits Administered as Poll Items

Table S2: Availability of Demographic Information for Each Poll

Table S3: Mean Effect Sizes for Responding Same Versus Different

Figure S1: Regression Coefficients for Change Over Time in Responding Same versus Different

Figure S2: Alternative Analysis That Assigned Numeric Values to Each Response Option

Figure S3: Demographic Subgroup Differences for Overall Stereotype Means

Figure S4: Item-level Data for the Most Recent Poll Conducted in April 2018

**Uploaded, on the Open Science Framework website (https://osf.io/kmwg8/register/564d31db8c5e4a7c9694b2be)**

data.zip: Data files needed for effect size generation and analyses

demoAnal.R: Code for demographic subgroup analyses

genES.R: Code for generating effect sizes

Figure1.R: Code for Figure 1

Figure2.R: Code for Figures 2 & S1

Figure3.R: Code for Figure 3

Figure4.R: Code for Figures 4 & S3

FigureS2.R: Code for Figure S2

FigureS4.R: Code for Figure S4

mainText.R: Code for main analyses, Tables 2 & S2

**Appendix S1: References to Included Opinion Polls**

CBS News. (2015). *CBS News Poll* [USCBS.021515.R12]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Eagly, A. H., Nater, C., Miller, D. I., & Sczesny, S. (2018). *New poll on gender stereotypes conducted by GfK: Government and academic omnibus panel.* Open Science Framework website: https://osf.io/kmwg8/register/564d31db8c5e4a7c9694b2be

Fox News. (2006). *Opinion Dynamics Poll* [USODFOX.092806.R39]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Gallup Organization. (1946). *Gallup Poll #370* [USAIPO1946-0370]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Gallup Organization. (1950). *Gallup Poll #452* [USAIPO1950-0452]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Gallup Organization. (1989). *Gallup News Service Poll: The women’s movement* [USAIPOGNS1989-89141-W2]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Gallup Organization. (1995). *Gallup/CNN/USA Today Poll: Drugs* [USAIPOCNUS1995-9508011]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Gallup Organization. (2000). *December Wave 1* [USAIPOGNS2000-53]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

New York Times. (1983). *New York Times poll: Women’s Survey* [USNYT1983-WOMEN]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Pew Research Center. (2008). *Pew social trends: Gender.* Retrieved from www.pewsocialtrends.org/2009/10/23/gender-data

Pew Research Center. (2015). *Gender and leadership omnibus.* Retrieved from www.pewsocialtrends.org/2016/05/25/gender-and-leadership-omnibus

Roper Organization. (1946). *Men and women* [USRFOR1946-054]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Roper Organization. (1977). *Roper Reports #77-2* [USRPRR1977-02]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Roper Organization. (1985). *Roper Reports #85-9* [USRPRR1985-09]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

U.S. News and World Report. (1987). *Attitudes* *Poll* [USRSPUSN1987-744-063]. Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

Virginia Slims. (1974). *The 1974 Virginia Slims American Women’s Opinion Poll* (Virginia Slims Poll #1974-0547) [USRSPVASLIMS1974-0547]*.* Retrieved from https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll

**Appendix S2: Analytical Strategy for Addressing Effect Size Dependencies**

Some polls administered multiple items for each stereotype domain (e.g., *creativity* and *intelligence*, both pertaining to competence), creating dependent effect sizes for our analyses. Our preregistration protocol detailed how we addressed these effect size dependencies (see https://osf.io/kmwg8/register/564d31db8c5e4a7c9694b2be). As explained in that document, we hoped to use robust variance estimation (RVE) but were mindful that RVE can perform poorly when the Satterthwaite degrees of freedom are less than four, which is a relevant consideration for our study given the small number of polls (Tanner-Smith et al., 2016; Tipton, 2015). We therefore decided to conduct a preliminary analysis to judge the applicability of RVE by examining simple regression models that used poll year to predict our three focal outcomes: the log odds of selecting women versus men for agency, communion, and competence items. Our preregistered decision rule was to judge RVE as inappropriate for our purposes if the Satterthwaite degrees of freedom for poll year were less than four for at least two of these three regression models.

After preregistering this analytic decision, we found that the Satterthwaite degrees of freedom for poll year were indeed small for the models of the agency (*df* = 3.46), communion (*df* = 2.36), and competence (*df* = 3.50). We therefore decided to instead use our back-up method of averaging effect sizes within polls for each of the three stereotype domains. For instance, effect sizes for Gallup (1995)’s *creative* and *intelligent* items were averaged to compute a composite effect size for the competence dimension. In this way, dependent effect sizes were handled at the effect size computation, rather than analysis, stage. We used the *survey* package in R to compute these averaged effect sizes while simultaneously accounting for (a) survey weights and (b) nesting of responses within participants (Lumley, 2017). These averaged effect sizes were then analyzed using the methods described in the main text (i.e., mixed-effects meta-regression models estimated using restricted maximum likelihood estimation and the Knapp-Hartung adjustment). However, as also noted in our preregistration, we nevertheless still used RVE for our analyses of items’ evaluative content because those analyses could not aggregate effect sizes by domain given that the item-level differences were the focus of those analyses.

The main text reports findings for effect sizes averaged by domain, but we nevertheless examined results for RVE models of item-level data as a sensitivity analysis of model specification. Consistent with the analyses reported in the main text, meta-regression models implementing RVE found that the odds of ascribing traits to women versus men significantly increased over time for communion items (*b* = 0.042; *SE* = .007; *p* = .018; *df* = 2.36) and competence items (*b* = 0.021; *SE* = .004; *p* = .009; *df* = 3.50), but not agency items (*b* = -0.010; *SE* = .008; *p* = .280; *df* = 3.46). Hence, our central results about historical change did not substantially differ using RVE versus averaged effect sizes.

**Appendix S3: Additional Details and Analyses Regarding the Equal Response Category**

This supplemental appendix details considerations and analyses about interpreting responses that the sexes are the same (i.e., equal alternative responses). Analyses reported in the main text instead focused on directly comparing the frequencies of women are more versus men are more responses. Analyses addressed the equal alternative response option in three ways: (a) comparing responding same versus different, (b) conducting multinomial analyses, and (c) analyzing mean stereotype scores based on assigning numeric values to each response option.

**Response structure for indicating same or equal.** For all items in all included polls, at least one respondent provided a response indicating that the sexes are the same or equal (unweighted average across items = 31%; min = 6%; max = 86%). This type of responding was common even for polls that did not explicitly mention the option in the question format. For instance, in the phone-based Fox News (2006) poll, respondents were asked, “do you think men or women are more intelligent?” Even though the pollster did not mention an equal alternative option, most respondents (59%) volunteered such a response. In all 10 polls that did not explicitly provide an equal alternative option, respondents could always volunteer this response to the pollster because those 10 polls were conducted over the phone or in person. No included poll therefore had a two-alternative forced choice response structure. This detail made it important to conduct supplemental analyses to study patterns of equal alternative responding.

**Comparing responding same versus different.** As shown in Figure S1 (panel a), compared to responding that the sexes are different, responding that the sexes are the same did not significantly change over time for agency or communion, simplifying interpretation about change over time for agency and communion. However, equal alternative responding significantly increased over time for competence and intelligence, especially for the data in the 1980s and onwards. This result especially indicated the need to conduct multinomial analyses to study how all response categories changed relative to each other.

**Multinomial analyses.** Figure 3 in the main text reported the results from multinomial logistic regression analyses to show how the different response categories changed relative to each other. Specifically, this analysis implemented Begg and Gray’s (1984) method for estimating multinomial models based on a set of binomial regressions comparing a reference category (in this case, the equal category) to others (women more, men more, or no response). We used Begg and Gray’s approach because, to our knowledge, other ways of estimating multinomial models have not yet been developed for meta-analytic models (see the R code in Figure3.R for details on how these models were implemented and combined to plot trend lines in Figure 3). Panels b and c of Figure S1 provides additional results from these multinomial models by showing how women are more and men are more responses changed relative to the equal response category. As shown in panel c, the change over time in competence and intelligence stereotypes was driven largely by decreases in men are more competent/intelligent responding, relative to equal alternative responding.

**Analyses based on assigning numeric values to each response option.** Supplemental analyses further addressed the equal alternative option by assigning numeric values to each response option: -1 = men more, 0 = same/equal, 1 = women more. These scores explicitly incorporate the percentage of same/equal responses into the stereotype magnitude because the mean stereotype score would be smaller in magnitude for larger frequencies of same/equal responses. As shown in Figure S2, results regarding change over time were similar with this alternative metric: Communion and competence increasingly favored women over time; agency showed no change.

Table S1

*Likability Ratings of the Traits Administered as Poll Items*

| Trait | Synonym (if necessary) | Mean likability | *SD* |  |
| --- | --- | --- | --- | --- |
| Communion  |  | 4.76 | 0.73 |  |
| honest |  | 5.57 |  |  |
| affectionate | warm-hearted | 5.32 |  |  |
| compassionate |  | 5.29 |  |  |
| patient |  | 5.13 |  |  |
| unselfish |  | 5.08 |  |  |
| polite and well-mannered | polite | 5.04 |  |  |
| outgoing |  | 4.81 |  |  |
| romantic |  | 4.65 |  |  |
| ability to handle people well | people person | 4.60 |  |  |
| sensitive |  | 3.84 |  |  |
| emotional |  | 3.07 |  |  |
| Agency  |  | 3.12 | 1.60 |  |
| ambitious |  | 5.01 |  |  |
| confident |  | 4.98 |  |  |
| courageous |  | 4.92 |  |  |
| hardworking | diligent | 4.74 |  |  |
| independent |  | 4.70 |  |  |
| ability to make decisions | decisive | 4.23 |  |  |
| decisive |  | 4.23 |  |  |
| proud |  | 3.90 |  |  |
| aggressive |  | 2.57 |  |  |
| critical |  | 2.13 |  |  |
| stubborn |  | 1.84 |  |  |
| strong | forceful | 1.79 |  |  |
| demanding |  | 1.63 |  |  |
| possessive |  | 1.35 |  |  |
| arrogant |  | 1.06 |  |  |
| selfish |  | 0.80 |  |  |
| Competence  |  | 4.67 | 0.43 |  |
| smart |  | 5.10 |  |  |
| creative |  | 5.06 |  |  |
| intelligent |  | 4.99 |  |  |
| common-sense | reasonable | 4.91 |  |  |
| willing to accept new ideas | broad-minded | 4.84 |  |  |
| level-headed |  | 4.80 |  |  |
| logical |  | 4.63 |  |  |
| ability to create or invent new things | inventive | 4.51 |  |  |
| innovative | inventive | 4.51 |  |  |
| thoroughness in handling details | thorough | 4.44 |  |  |
| organized | systematic | 3.56 |  |  |

*Note*. Mean likability ratings are from Chandler (2018) and are based on a 7-point scale ranging from 0 (*least favorable or desirable*) to 6 (*most favorable or desirable*). Chandler’s analyses determined that the stereotypical attribute “aggressive” had become more negative in likability over time. Therefore, we averaged the likability value for this item across the replications (Anderson, 1968 = 3.04, Dumas, Johnson, & Lynch, 2002 = 2.69, and Chandler, 2018 = 1.97). For one attribute (*calm in emergencies*), no match was possible.

Table S2

*Availability of Demographic Information for Each Poll*

|  |  |
| --- | --- |
|  | Demographic variable |
| Poll | Sex | Age | College education | Race/ ethnicity | Marital status | Employment status |
| Roper (1946) |  |  |  |  |  |  |
| Gallup (1946) |  |  |  |  |  |  |
| Gallup (1950) |  |  |  |  |  |  |
| Virginia Slims (1974) |  |  |  |  |  |  |
| Roper (1977) |  |  |  |  |  |  |
| New York Times (1983) |  |  |  |  |  |  |
| Roper (1985) |  |  |  |  |  |  |
| U.S. News (1987) |  |  |  |  |  |  |
| Gallup (1990) |  |  |  |  |  |  |
| Gallup (1995) |  |  |  |  |  |  |
| Gallup (2000) |  |  |  |  |  |  |
| Fox News (2006) |  |  |  |  |  |  |
| Pew (2008) |  |  |  |  |  |  |
| Pew (2015) |  |  |  |  |  |  |
| CBS News (2015) |  |  |  |  |  |  |
| Eagly et al. (2018) |  |  |  |  |  |  |

*Note*. Shaded cells indicate which polls had data for each demographic variable. The raw data for Fox News (2006) were unavailable and therefore that poll was not included in any demographic subgroup analyses, as indicated in the Method section on p. 12.

Table S3

*Mean Effect Sizes for Responding Same Versus Different*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Stereotype measure | *k* | Mean | 90% predictioninterval | *τ2* | *I2* |
| Communion | 12 | 24% | 8%, 54% | 0.63 | 99.61% |
| Agency | 9 | 33% | 13%, 62% | 0.52 | 99.59% |
| Competence | 11 | 39% | 15%, 71% | 0.65 | 99.70% |
|  Intelligence | 8 | 59% | 22%, 88% | 0.98 | 99.61% |

*Note*. *k* = number of polls; mean= random-effects weighted mean of the percentage of respondents indicating that women and men are the same or equal rather than different; prediction interval = the middle 90% of the true underlying effects; *τ2 =* tau-squared, the estimated between-poll variance of effect sizes on a log odds scale; *I2 =* percentage of total variability in effect sizes due to true between-poll heterogeneity rather than chance*.* Heterogeneity was significant for all stereotype measures (all *p*s < .001).

\* *p* < .05. \*\**p* < .01. \*\*\**p* < .001



*Figure S1*. Regression coefficients and 95% CIs for change over historical time in responding equal/same versus different. Positive values indicate an increase over time in the odds of indicating that (a) women and men are equal (as opposed to different), (b) women are more (as opposed to equal), and (c) men are more (as opposed to equal). The bars (from left to right) indicate regression coefficients without controls, with controls for each of two potential confounds, and with the removal of data from the 1940s and 1950s.



Figure S2. Alternative analysis that assigned numeric values to each response option. These values were: -1 = men more, 0 = same/equal, 1 = women more. Each data point represents the mean stereotype score for each poll based on these numeric values with the 95% CIs for (a) communion, (b) agency, (c) competence, and (d) intelligence.

**

*Figure S3.* Demographic subgroup differences for overall stereotype means. Values indicate the percentages of “women more” responses among sex-differentiated responses (i.e., “women more” or “men more” responses). The values differ from those in Figure 4 in main text because this figure shows meta-analytic averages across all poll years, whereas Figure 4 shows predictions for the year 2018. The dot colors indicate that participants significantly ascribed traits more to women (red) or men (blue) or that their responses did not significantly favor either sex (gray).



*Figure S4*. Item-level data for the most recent poll conducted in April 2018. Values indicate the percentages of respondents who ascribed the trait more to men or women among all respondents and do not sum to 100% because of respondents who indicated that women and men are the same. In addition, the percentages of respondents who did not provide a response are not shown, but those percentages were less than 1% for all items.