

The Global Decline in Sexism: A Multilevel Meta-Analytic Review of Trends in Countries' Hostile Sexism, Benevolent Sexism, and Gender Inequality Over Time

Participant-visible method context. Results, discussion, captions, and the target figure are excluded.

Transparency and Openness We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidance (Page et al., 2021) and the Meta-Analysis Reporting Standards guidelines for reporting metaanalyses (Appelbaum et al., 2018). Our research goals, search strategy, and inclusion/exclusion criteria were preregistered. The preregistration, data, analysis code, and research materials are available on the Open Science Framework (Hammond & Karl, 2025). Differences from the preregistration are summarized in Table 1.

Search Strategy The first and second authors searched for the term “ambivalent sexism inventory” (ASI) in publications from 1996 to 2023 (inclusive) in Google Scholar, ProQuest, ProQuest Dissertations, Scopus, and Web of Science.¹ Search and screening waves occurred

in August–December 2022 and May–June 2024. This search term identified any available study—including papers/theses, technical reports, preprints, and conference reports—that either measured the ASI and/or contained an English-language citation of Glick and Fiske (1996) published on or before December 31, 2023 (including online publication ahead of print). Samples were included if they reported the mean and standard deviation of hostile sexism and/or benevolent sexism and/or the correlation between hostile sexism and benevolent sexism. Samples were excluded if (a) the ASI measurement occurred after an experimental manipulation or intervention; (b) the target sample was people with diverse sexualities, for which the ASI has not been validated (see Cross et al., 2021; Glick, 2023); or (c) the target sample was people who were violent offenders in treatment.

Data Screening Our search followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses model (Figure 1). Most reports matching the search term did not pass the screening stage (e.g., abstracts indicated a qualitative study, a review, or legal scholarship). Of the reports assessed for eligibility, the highest number of exclusions was coded as “unavailable” due to lacking necessary information for the meta-analysis (i.e., descriptive statistics or information on scale points). In some of those cases, data were unavailable because the ASI was included as a filler task or footnoted as a supplementary test. However, many reports failed to make descriptive statistics available as required by APA reporting standards. Finally, some reports cited the ASI for another purpose, including qualitative research or adaptations of the ASI for other purposes (e.g., indexing experiences of sexism, developing vignettes of sexism, assessing sexism toward specific social groups), and were coded as ASI “not measured.”

Data Extraction Data were extracted by the first and second authors by recording means, standard deviations, and correlations on an Excel spreadsheet for hostile sexism and benevolent sexism. We also recorded characteristics of the measures (i.e., language, number of items, possible scale range, reliabilities) and samples (i.e., sample country, student sample, online sample). For all papers that indicated supplementary materials or open data, we also searched those sources to supply the required information. Data on the country-level indices of inequality for years up to and including 2022 (i.e., data published in the most recently available report; United Nations Development Programme, 2024) were obtained from the United Nations Human Development data center (United Nations Development Programme, n.d.): the

The Web of Science search comprised the following database subscriptions: Arts and Humanities Citation Index, Conference Proceedings Citation Index, Conference Proceedings Citation Index—Social Sciences and Humanities Edition, Emerging Sources Citation Index, MEDLINE, Social Sciences Citation Index, Book Citation Index (Science + Social Sciences), KCI-Korean Journal Database, Russian Science Citation Index, SciELO Citation Index, BIOSIS Citation Index, Chemical Collection (Index Chemicus + Current Chemical Reactions), Current Contents Connect Collection, Book Citation Index (Science + Social Sciences), CAB Abstracts On CABI, and Science Citation Index Expanded.

HAMMOND, SINGH, AND KARL

Table 1 Summary of Changes Following the Preregistration Preregistration plan Method: Grand-mean centering variables. Search strategy: Timeframe of data from “1996–2021.” Search strategy: Databases.

Change

Reason

Year remains centered on the midpoint. All other predictors were z-scored. Extended to “1996–2023.”

Data extraction: Inequality indices extracted from United Nations reports.

Now also includes manuscripts and data in <https://osf.io/> and PsyArXiv. Data obtained from the United Nations Human Development data center.

Analysis: Latent growth curve models assessing associations between trajectories.

Multilevel meta-analysis models estimating year as a random effect.

Primary Analysis 2: Regressing sexism on indicators on inequality.

Adding a test predicting the sample-level differences between men’s and women’s endorsement of benevolent sexism. Replaced with the United Nations scores of countries’ GII.

Primary Analysis 3: Assessing United Nations scores of countries’ GEM.

Required to reach convergence in the random effects models. Using the latest full-year data at the time of article revision, recommended in the peer-review process. Expansion of search recommended in peer review. The data center was made accessible after the date of preregistration; indices were available for all years, rather than exclusively for years of published reports. Growth curve models failed to converge. We opted for an alternative strategy recommended in peer review: multilevel meta-analysis including random effects for intercepts and slopes (where possible). This analysis tests the “Protection Racket” as described in Glick et al. (2000), recommended in peer review. The GII was introduced in 2010 as a measure of gender inequality that superseded the GEM.

Note. GII = Gender Inequality Index; GEM = Gender Empowerment Measure.

Gender Inequality Index (GII), Gender Development Index (GDI), and Human Development Index (HDI).

Reliability Due to the large number of included studies, we assessed the reliability of data extraction by double-coding extracted data from a random selection of 50 included samples. Out of 1,030 entered points of data for those samples, coders had perfect agreement on 982 (95.3%). Data on the means, standard deviations, country, and sample type had perfect agreement. Coder disagreement occurred in 19 cases: scale reliabilities ($N = 4$; <1%), sample size ($N = 7$; <1%), number of scale items ($N = 4$; <1%), and scale range ($N = 4$; <1%). When checking the original sources, we identified that two studies had provided different information between their text and tables (subsequently marked “data unavailable” and excluded). In 29 cases, only one coder found information on correlations ($N = 8$; <1%), scale reliabilities ($N = 19$; 2%), or sample size ($N = 2$; <1%); these were confirmed by checking the original sources. Next, we checked for errors by examining for unexpectedly high, low, or missing values in our dataframe. Fifteen values were checked against their original sources, nine of which were entry errors and one was a missed value. Finally, we manually examined our dataframe for duplicated samples across different studies (e.g., a thesis subsequently published in a journal, reanalysis of samples in different publications). In 55 identified cases of duplication, we included only the earliest publication (see Figure 1).

Sample Characteristics The final dataframe comprised 191,265 women and 148,476 men gathered in 1,097 samples from 81 countries. However, some studies exclusively examined women or men or, similarly, exclusively hostile sexism or benevolent sexism, and so we supply the N s for

samples and countries in each analysis. Nonetheless, our primary analyses represented a large proportion of the total dataframe (e.g., 133,879 women and 103,425 men for estimation of hostile sexism over years). In the Supplemental Materials, we considered potential differences in endorsement of sexism depending on characteristics of the sample (i.e., university, online, or community sample) or measure (i.e., full vs. short-form versions of the ASI), and report similar patterns of results to those reported below.

Convergent Validity We conducted a test to establish convergent evidence that our meta-analytic estimates aligned with an external measure of gender attitudes. Given the lack of any representative cross-cultural and longitudinal data on ambivalent sexism, we selected Brandt's (2011) proxy of hostile sexism from the WVS, which averaged the items "On the whole, men make better political leaders than women do" and "On the whole, men make better business executives than women do." These items take a focus on suitability for high-status gender roles and thus do not directly index hostile sexism, which represents women as underhandedly attempting to emasculate and disempower men (see Glick & Fiske, 1996). Accordingly, although we did not expect strong associations with our estimate of hostile sexism, positive associations should emerge between the WVS proxy and our measure of hostile sexism as indicators of negative evaluations toward women. We accounted for missingness in the WVS with Multiple Imputation by Chained Equations, extracting 100 imputed data sets with 100 iterations, with the mice package (Version 3.16.0; Van Buuren & Groothuis-Oudshoorn, 2011) in R (Version 4.3.0). We then estimated the associations between the aggregate measures of hostile sexism with the WVS proxy for hostile sexism, matching the same country and year, for women and men. Significant small-to-moderate associations emerged between hostile sexism and the WVS proxy scores for women ($r = .18$,

A META-ANALYTIC REVIEW OF SEXISM AND INEQUALITY

Included

Screening

Identification

Figure 1 Flow Diagram of Data Identification, Screening, and Inclusion for Our Meta-Analysis of Measures of ASI Since Its First Publication (1996–2023) Database records identified: Google Scholar ($n = 7494$) ProQuest ($n = 2928$) Scopus ($n = 2939$) Thesis Database ($n = 1051$) Web of Science ($n = 177$) Preprints ($n = 15$)

Duplicates Removed ($n = 6360$)

Records screened (abstracts) ($n = 8243$)

Records excluded ($n = 6470$)

Reports sought for retrieval ($n = 1773$)

Reports not retrieved: Manuscript unavailable ($n = 28$)

Reports assessed for eligibility ($n = 1745$)

Reports excluded: Statistics not reported ($n = 373$) ASI not measured ($n = 252$) Experiment/Intervention ($n = 211$) Aggregated ASI subscales ($n = 89$) Duplicated samples ($n = 58$) Violence Intervention ($n = 21$) Non-heterosexual sample ($n = 12$)

Total samples ($k = 1,097$ from 729 reports) Total participants ($N = 339,740$)

Note. ASI = Ambivalent Sexism Inventory.

unstandardized $B = .30$, 95% CI [0.28, 0.32], $t = 35.55$, $p < .001$) and for men ($r = .25$, unstandardized $B = .39$, 95% CI [0.38, 0.41], $t = 44.65$, $p < .001$). Thus, supporting convergent validity, our metaanalytic estimation of countries' hostile sexism covaried with a proxy for hostile sexism obtained from representative samples within those countries.