Sex Roles (2025) 91:21
https://doi.org/10.1007/s11199-025-01561-x

ORIGINAL RESEARCH

The Fragility of Scientific Knowledge: A Case Study on the Miscitation of Findings on Gender Stereotypes

Christa Nater[1] · **Alice H. Eagly**[2]

Accepted: 23 January 2025

Abstract

Miscitation of research findings is a common problem as evidenced by 19% of citations in top psychology journals being in error (Cobb et al., American Psychologist, 79:299–311, 2024). Such errors interfere with the orderly cumulation of knowledge. Providing a case study, this research examines the citations of a recent and highly cited article on gender stereotypes (Eagly et al., American Psychologist, 75:301–315, 2020), which found that communion, agency, and competence stereotypes each showed a distinctive trend over time. Analysis of the 751 documents that cited this article's findings showed that overall, 59% of citations were accurate and 9% somewhat accurate, yet a surprisingly high rate of 32% were inaccurate. These inaccuracies most often misrepresented findings on agency with 37% of the citing articles being inaccurate, and among these inaccurate citations, 21% directly contradicted the findings by erroneously stating that the tendency to ascribe agency more to men than women had faded over time. Miscitations for the two other stereotype domains were less egregious. Of the communion citations, 25% ignored that the tendency to ascribe communion more to women than men has become stronger over time. Of the competence citations, 18% ignored that most people in recent years believe that women and men are equally competent. The discussion considers possible reasons for misciting findings on gender stereotypes, particularly for the agency stereotype that has favored men over women ever since the 1940s. We further expound on the feminist theme of the fragility of scientific knowledge, especially when research findings compete with preconceptions that people, including researchers, may have about the phenomena of gender.

Keywords Gender stereotypes · Citation accuracy · Miscitation · Scientific writing · Stereotype change

In science, citations of research findings serve to describe the current state of knowledge on a topic and aid the interpretation of new research. Posing a threat to scientific integrity, however, the miscitation of findings is a common problem in research practice. As demonstrated by Cobb et al. (2024), approximately 19% of citations across 89 articles in eight of the top psychology journals were in error, by either failing to note important details of cited findings (9.3%) or by blatantly misdescribing them (9.5%). Similarly, an analysis of medical journals revealed that about 12% of the citations were seriously incorrect (Jergas & Baethge, 2015). Some instances of miscitations have become "academic urban legends" (Rekdal, 2014), such as the case in which a decimal point error in citing a finding apparently produced the popular belief that spinach is an excellent nutritional source of iron (Hamblin, 2010). Such citation errors can be passed along in academic and other writing, thus interfering with the orderly cumulation of valid findings and threatening the integrity of science.

This research project aims to elucidate the phenomenon of miscitation by close study of the citations of a single recent article on how gender stereotypes changed over time (Eagly et al., 2020). This article has received wide attention from both the public through expansive press coverage (e.g., Hoffower, 2019; Salam, 2019) and within science through numerous citations in scientific journals. Yet, our casual observations of inaccuracies—particularly in the agency stereotype—motivated our investigation of the type and extent of these miscitations, given that such errors interfere with the orderly accumulation of knowledge. This project

✉ Christa Nater
christa.nater@unibe.ch

[1] Department of Psychology, University of Bern, Bern, Switzerland

[2] Department of Psychology, Northwestern University, Evanston, IL, USA

Published online: 11 March 2025

thus provides a case study exploring the accuracy of citation when research findings must make their way through presumed preconceptions of the researchers who cite them and accurately or inaccurately convey their message into a body of knowledge on gender.

To provide context for these miscitations, we first describe the findings of the Eagly et al. (2020) study on gender stereotypes, which are defined as widely shared beliefs about the attributes or characteristics of women and men. These stereotypes can have far-reaching consequences because they influence the perception and evaluation of women and men in general and thereby constrain the possibilities of the individual members of these social categories (e.g., Schmader & Nater, in press). The Eagly et al. (2020) study reported changes in gender stereotypes, as assessed in nationally representative U.S. public opinion polls conducted over the 1946–2018 timespan. Consistent with common themes in research on gender stereotype content, the authors classified the traits presented in the polls into the three dimensions of *communion* (e.g., affectionate, emotional), *agency* (e.g., ambitious, courageous), and *competence* (e.g., intelligent, creative). Poll respondents had indicated whether each trait (e.g., affectionate, ambitious, intelligent) was "more true of women," "more true of men," or "equally true of both." Eagly et al. (2020) then meta-analyzed the data to test how U.S. respondents' beliefs about women and men had changed over seven decades.

As shown in Figs. 1 and 2, the key findings were that, for communion, women were viewed as increasingly more communal than men over time. Illustrating this trend, among respondents who indicated a difference between women and men, 54% thought women were more communal than men in 1946, 83% did so in 1989, and 97% in 2018 (see Eagly et al., 2020). For agency, men were seen as more agentic than women since the 1940s with no change over time. Finally, for competence, equality of women and men gained significantly over time to become the most highly endorsed response. Specifically, the percentage of respondents viewing women and men as equally competent was 26% in 1946 but 69% in 2018. Among those respondents who did not indicate equality, the 1940s belief that men are more competent reversed over time to become a small advantage for women more recently.

These findings document that each of these three components of gender stereotypes—communion, agency, and competence—showed a distinctive trend over time. Therefore, accurate representations of the findings must be tailored to each of these components of meaning and reported accurately, not compromised by beliefs that people, including researchers, may have about the stereotypical attributes of women and men.

Miscitations may have arisen because the study's findings competed with the citing authors' preconceptions about the phenomena of gender. Because gender is a prominent feature of daily life, certain ideas about gender may have gained considerable *cultural fluency* (Oyserman & Yan, 2019) in the sense of "feeling right," that is, reflecting common knowledge about how the world works. For

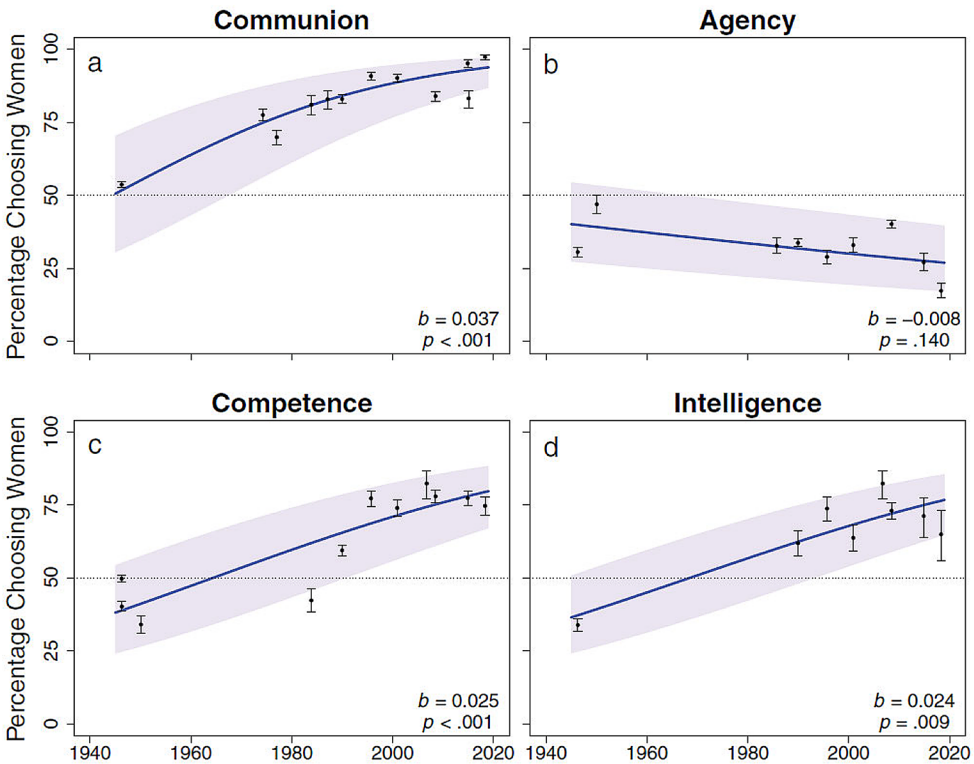

Fig. 1 Stereotype Change According to Respondents Choosing That Women More Than Men Possess Traits, Separated by the Stereotype Dimensions (as published in Eagly et al., 2020) *Note*. Change over historical time in the mean percentage (and 95% confidence intervals) of respondents choosing *more true of women* for communal (Panel **a**), agentic (Panel **b**), competent (Panel **c**), and intelligent (Panel **d**). These figures do not include respondents who chose *equally true for both*. Reproduced by permission of the American Psychological Association.

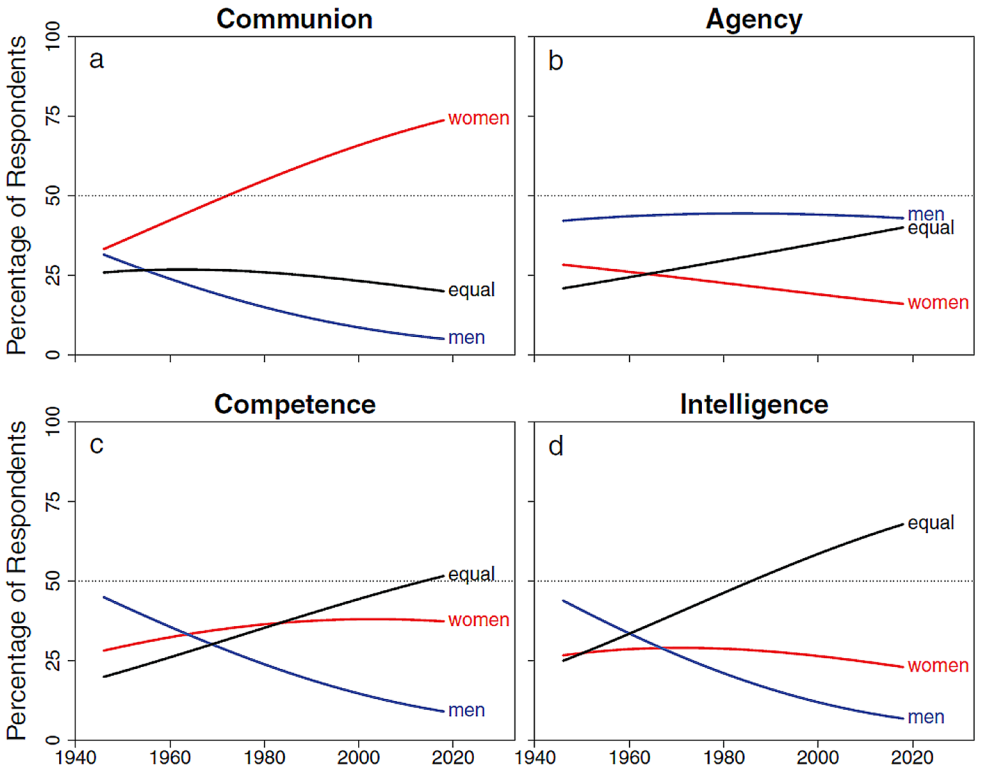

Fig. 2 Stereotype Change Including Respondents Choosing That a Trait is Equally True for Both Women and Men (as published in Eagly et al., 2020). *Note.* Stereotypes displaying change over historical time in the mean percentage of respondents indicating more true of women, more true of men, or equally true for both on communal (Panel **a**), agentic (Panel **b**), competent (Panel **c**), and intelligent (Panel **d**). Reproduced with permission of the American Psychological Association.

example, consistent with social role theory (Eagly & Wood, 2012; Koenig & Eagly, 2014), belief in increasing gender similarity may have become more culturally fluent as the social roles of women and men have become more similar due to increasing equality in labor force participation and education (e.g., Blau & Winkler, 2018). However, social psychologists have long provided the counternarrative that stereotypes of social groups are rigid and unchanging (see Hinton, 2020), suggesting that it may be culturally fluent to believe that gender stereotypes are stable over time. This narrative is also reflected in dictionary definitions stating that stereotypes are "Something continued or constantly repeated without change" (Oxford English Dictionary, n.d.). Finally, ideological bias might lead authors to ignore or distort ideologically unwelcome findings in the manner that feminists might miscite findings showing the strengthening of stereotypes, and antifeminists might miscite findings showing the weakening of stereotypes over time.

Each of these various narratives of change and stability would be consistent with aspects of the Eagly et al. (2020) findings: Belief in increasing greater equality thus found verification in the rise of women's perceived competent qualities relative to those of men, and belief in stability found verification in the constancy of men's perceived agentic advantage over women. However, the rise of the communion of women relative to men could be more puzzling, given its inconsistency with either rising gender equality or beliefs that stereotypes are rigid and unchanging.

Miscitations may also arise from career demands at research universities. As claimed by Bauerlein et al. (2010), researchers usually face strong pressure to publish to be competitive for scarce academic positions and resources as well as to gain recognition in their field, leading them to "cut corners" in their research and writing practices to maximize publications. Citations may then become inaccurate, given how time-consuming it can be to carefully read and evaluate publications relevant to one's own paper. In fact, reports have suggested that up to 80% of authors do not read the full text of articles they cite (Simkin & Roychowdhury, 2006).

The importance of correctly representing research findings on gender, as on other topics, motivated our systemic investigation of the contours of the miscitations of the Eagly et al. (2020) article. We therefore retrieved all citations to the article and classified them by their degree of accuracy and other attributes. Using the Cobb et al. (2024) finding of 19% inaccuracy as a baseline, this project provides a case study of citation accuracy specific to knowledge of gender stereotypes. To better describe when and where miscitations occur, we explored whether accuracy related to the citing scientific discipline (e.g., psychology, medicine, gender studies, computer science), type of report (e.g., journal articles, dissertations, conference contributions), or journal metrics (i.e., impact factor). Finally, we examined author characteristics (i.e., gender and seniority of the first and last author; without expecting to find any effects, see preregistration).

Method

Transparency, Openness, and Reproducibility

The list of citing records, our coding of the citations, and the analysis code are available on OSF (https://osf.io/sfkp2). The preregistration of the training procedure for coders, the coding scheme, and the analysis strategy were uploaded before the start of the coding (https://osf.io/dy5tf). Data were analyzed using R, version 4.1.2.

Selection of Citations

Forward reference searches identified all published reports that cited Eagly et al. (2020) and were published until November 2023. As preregistered, this search was performed on Google Scholar (revealing $N = 751$ citing articles), Web of Science ($N = 383$), and Scopus ($N = 424$). On Google Scholar, the most comprehensive database, the search revealed journal articles ($N = 456$), dissertations ($N = 112$), books ($N = 56$), working papers ($N = 24$), preprints ($N = 20$), conference publications ($N = 10$), master's theses ($N = 45$), bachelor's theses ($N = 18$), and other documents ($N = 10$; e.g., encyclopedia entry, blog posts). The non-English reports ($N = 67$) were translated into English using machine-based translators (e.g., DeepL; miscitations were not more common in these translated reports; for more details, see Supplement A in the online supplement).

Coding Procedures

Coding Scheme

Miscitation varies by the degree to which a report has provided a correct and complete account of the relevant findings. Absent a validated scale for indexing the degree of miscitation, we developed a coding scheme based in part on Cobb et al. (2024). The coding pertained to three gender stereotype domains, namely, communion, agency, and competence. An additional category coded for citations not referring to only one of these dimensions: not one gender stereotype domain specifically.

As preregistered, the coding scheme operationalized citation accuracy into the following four broad categories: (1) accurate, (2) somewhat accurate, and (3) inaccurate (see Table 1). In addition, citations not pertaining to any findings and thus disallowing evaluation of accuracy were coded as (0) definitions only, such as definitions of gender stereotypes or references to social role theory. Given that these "definition only" citations did not report findings, we excluded them from our analyses.

Table 1 displays the accuracy coding categories: *1-Accurate* indicated the correct description of the findings of Eagly et al. (2020) article without omission of any key aspect; *2-Somewhat accurate* included two subcategories: *2a-Omission of communion increase* signified omission of the finding that belief in women as communal has increased over time and instead presented it as stable; *2b-Omission of competence equality* signified omission of the finding that competence was equally associated with men and women and instead presented it as greater female competence. *3-Inaccurate* included four subcategories: *3a-Overspecification to prescriptions* signified the application of the article's findings regarding descriptive stereotypes to prescriptive stereotypes, which were not reported in the article; *3b-Generalization* signified the generalization of the findings to unassessed processes and outcome variables (e.g., self-perception, self-ratings) or social groups (e.g. children, race/ethnicity); *3c-Aspect not included in study* signified that the cited gender stereotype finding(s) were not assessed in the article; *3d-Contradiction* signified statements about gender stereotypes directly contradicting the article's results.

Coding of Additional Report Characteristics

One coder retrieved objective information about each report: author names, publication year, type of publication (i.e., journal article, dissertation, master's thesis, bachelor's thesis, book, conference publication, preprint, working paper, other), journal's impact factor and H-index (www.resurchify.com/if/impact-factor-search), the journal's scientific field (www.scimagojr.com/journalsearch.php), language of the report, and number of authors. For the first and the last author, we additionally coded the gender (i.e., woman, man, diverse, unknown) and seniority (i.e., student, postdoc, assistant professor, associate professor, full professor, practice or industry, unknown). To identify gender, the coder used the first name and examined the author's departmental website for other evidence (e.g., pronouns). To identify seniority, the coder visited authors' departmental websites, where staff ranks are listed.

Training of Coders and Coding

The team of coders consisted of three research assistants. As part of their training, each coder carefully studied the article by Eagly et al. (2020) and wrote a detailed summary of its methods and findings. As preregistered, in a meeting with the first author, they then discussed their summaries to ensure each coder had an accurate understanding. Each coder then coded three training articles, followed by a discussion with the first author and the other coders. Subsequently, each coder independently coded the first 30 citing

Table 1 Coding Categories with Examples

Coding category	Examples			
	Communion	Agency	Competence	Not one specifically
0-Definition only	Beliefs that women are more communal.	Beliefs that men are more agentic.	Beliefs that men and women are equally competent.	Gender stereotypes refer to societally held ideas about what people of a particular gender are like.
1-Accurate	The belief that women are more communal than men has increased since the 1940s.	The belief that men are more agentic than women has remained stable over time.	The belief that men and women are equally competent has increased over the last decades.	Gender stereotypes have changed over the past 70 years.
2a-Omission of communion increase	Since the 1940s, women are still believed to be more communal than men.	-	-	-
2b-Omission of competence equality	-	-	Women are increasingly believed to be more competent than men.	-
3a-Over-specification to prescriptions	The belief that women should be communal has increased.	The belief that men should be agentic has remained stable.	The belief that women and men should be equally competent has increased.	Society holds certain prescriptive stereotypes about how people of different genders are expected to act.
3b-Generalization	Young girls are increasingly believed to be more communal.	Men rate themselves higher on agency.	Stereotypes surrounding a mother's competency have improved.	Women's development in working life may, for example, have caused them to attribute themselves emotions that men previously had a monopoly on.
3c-Aspect not included in study	Women show increasingly communal behavior.	In the future, men might be believed to be less agentic.	Women's increase in competence makes them more suitable as leaders.	Women are disadvantaged in taking better paid and socially respected positions.
3d-Contradiction	Women's communion has remained stable since the 1940s.	Men and women are now perceived as similarly agentic.	In general, women are believed to lack competence.	Stereotypes tend to minimalize, but they have not vanished and probably will not.

reports. These reports were also consensus-coded by the two authors. A second meeting then served to compare each coder's results with those of the two authors. Any deviation was discussed, and misconceptions were corrected.

From the team of three coders, a subset of two coders independently coded each citation. Weekly meetings with the first author ensured that the coders did not drift apart over time and secured high-quality coding. In these meetings, we further addressed questions and discussed coding disagreements to reach a consensus. Coders had access to the verbatim citations assigned to them but no access to the other coders' coding or the additional report characteristics.

Interrater Reliability

The raters' agreement was high as assessed by the percentage of agreeing responses relative to the total number of codings. For the main categories, agreement was 93% for accurate, 83% for somewhat accurate, and 89% for inaccurate. Given that the agreement for each dimension was above the preregistered threshold of 70%, no additional coders were added.

An additional interrater reliability statistic consisting of a mean-rating ($k=2$), consistency-agreement, 2-way random effects model found high reliability within each stereotype

Table 2 Citation Accuracy Across and Within Each Stereotype Domain

Coding category	Overall	Communion	Agency	Competence	Not one specifically
Main categories (total 100%)					
1-Accurate	59%	52%	63%	71%	52%
2-Somewhat accurate	9%	25%	-	18%	-
3-Inaccurate	32%	23%	37%	11%	48%
Separate for subcategories (total 100%)					
1-Accurate	59%	52%	63%	71%	52%
2a-Omisson of communion increase	5%	25%	-	-	-
2b-Omission of competence equality	4%	-	-	18%	-
3a-Overspecification to prescriptions	3%	7%	6%	0%	1%
3b-Generalization	2%	1%	3%	2%	1%
3c-Aspect not included in study	19%	10%	8%	5%	40%
3d-Contradiction	8%	4%	21%	3%	6%

Note. Overall consists of the following four categories: communion (N=204 citations), agency (N=198 citations), competence (N=237 citations), and not one specifically (N=356 citations). For the main category 2-Somewhat accurate, we preregistered an additional subcategory called "double barreled," which designated citations that "entailed two claims within a single statement, whereas the original study only supported one of those claims" (following Cobb et al., 2024, p. 304). During the coding process, we eliminated this category and coded the respective statements in one of the more specific categories. Percentages in some columns may not total 100 due to rounding.

dimension: For communion, ICC = 0.91, 95% CI [0.89, 0.93], agency, ICC = 0.96, 95% CI [0.95, 0.97], competence, ICC = 0.94, 95% CI [0.93, 0.96], and not one specifically, ICC = 0.93, 95% CI [0.91, 0.94].

Analysis Procedure

The main outcome measure was the degree of citation accuracy (in %), both across and within each stereotype dimension. To explore inaccuracies, further analyses probed the content of the *2-Somewhat accurate* and *3-Inaccurate* codes (see preregistration).

As preregistered, analyses then focused only on journal articles to examine whether articles published in certain journals were more accurate, such as those with higher impact factors. This analysis included journals with at least three citations. Analyses then examined accuracy for the 20 journals that included the largest number of citing articles.

Exploratory analyses further examined whether accuracy related to (a) the first or last author's gender or seniority or (b) the citing report's scientific discipline (e.g., gender studies, psychology, medicine) or (c) the type of report (e.g., journal article, dissertation, working paper).

Results

Citation Accuracy Overall

First, 22% of the citations referred to the Eagly et al. (2020) article for definitional purposes only, were therefore coded as (0) definitions only, and were thus excluded from all subsequent analyses examining accuracy.

Table 2 displays the extent to which the article's findings were miscited. The first column displays the results across the three stereotype dimensions, the second to fourth columns separately display each stereotype dimension, and the last column displays the category *not one specifically*. As can be seen in Table 2 (top half), across the three stereotype domains, 59% of citations were accurate, 9% somewhat accurate, and 32% inaccurate (see column overall). For *communion*, 52% of citations were accurate, 25% somewhat accurate and 23% inaccurate. For *agency*, 63% were accurate, none (0%) was somewhat accurate, and 37% were inaccurate. For *competence*, 71% were accurate, 18% somewhat accurate, and 11% inaccurate. For the nonspecific category *not one specifically*, 52% were accurate, none (0%) was somewhat accurate, and a high 48% were inaccurate.

Regarding the reasons for inaccuracy (see subcategories listed in Table 2, bottom half), results indicated that the highest percentage of accurate citations occurred for competence (66%), followed by agency (40%) and communion (33%). The most severe form of miscitation, that is, directly contradicting the findings of stereotype change over time (i.e., 3d-Contradiction), pertained to agency (21%). Fewer citations contradicted the change for the other two stereotype dimensions, communion (4%) and competence (3%). Yet, a substantial percentage of citations were inaccurate also for these two domains by either omitting the finding that communion for women increased over time (25%) or that competence equality became more common (18%). Finally, regarding citations that did not directly refer to a stereotype dimension (i.e., coded as *not one specifically*), results found that a substantial 37% of all citations referred to aspects that were not examined in the Eagly et al. (2020) article.

Citation Accuracy Depending on the Journal

As preregistered, exploratory analyses examined whether a journal's impact factor related to the extent of miscitation. These analyses examined the percentages of citations across the three stereotype domains and the category *not one specifically*. Results showed that impact factor was not related to classification as accurate, $r(199) = 0.05$, $p = .499$, somewhat accurate, $r(199) = -0.07$, $p = .300$, or inaccurate, $r(199) = 0.05$, $p = .499$. Results further found that the impact factor was also not related to accuracy in the subcategories (see Supplement B, Table B1 in the online supplement).

An additional exploratory analysis examined accuracy in the 20 journals that included the largest number of citing articles. Results indicated that the Eagly et al. (2020) findings were most often cited in *Frontiers in Psychology* ($N = 21$), followed by *Sex Roles* ($N = 13$; see Table B2 in Supplement B in the online supplement). Accuracy results indicated that across all three stereotype dimensions, among the 20 most frequent journals (with four to 21 citations per journal), the highest percentage of accurate citations were in *Gender in Management*,*International Journal of Environmental Research*, and *Public Health* (i.e., 100% accurate). In contrast, citations most often contradicted the actual findings in *Journal of Applied Psychology* and *Group Processes & Intergroup Relations* (i.e., 3d-Contradiction, 100% inaccuracy; see Table B3 in Supplement B in the online supplement).

Citation Accuracy Depending on Gender and Seniority of First and Last Author

Exploratory analysis further examined the relations between the degree of miscitation and author characteristics. Table 3 displays the percentages for the main categories and subcategories separately by first author and last author gender and seniority.

The first set of chi-square tests focused on the first author's gender and seniority and the second set focused on the last author's gender and seniority. Consistent with our preregistered expectation, results showed no significant associations between the degree of miscitation and either the first author's gender, $\chi^2(2) = 0.42$, $p = .810$ or seniority, $\chi^2(10) = 14.12$, $p = .167$. Similarly, for the last author, results showed no significant associations for either gender, $\chi^2(2) = 0.54$, $p = .763$, or seniority, $\chi^2(10) = 2.50$, $p = .991$. These analyses excluded the categories *diverse* for gender and *unknown* for gender and author seniority, due to either the small number of citations (for diverse gender) or the lack of meaningful interpretation (for the category unknown).

Table 3 Accuracy of Coding Separated by the Gender and Seniority of the First and Last Author

	First author										Last author									
	Gender (%)			Seniority (%)							Gender (%)			Seniority (%)						
	W	M	uk	S	PD	AsP	AcP	FP	PI	uk	W	M	uk	S	PD	AsP	AcP	FP	PI	uk
Main categories (total 100%)																				
1-Accurate	62	63	64	57	74	62	64	65	66	57	64	68	52	66	61	60	65	68	57	65
2-Somewhat accurate	15	13	8	14	14	13	19	14	17	10	16	14	13	13	19	17	15	14	17	12
3-Inaccurate	22	24	28	29	13	25	17	20	17	33	20	18	35	21	19	23	20	19	27	23
Separate for subcategories (total 100%)																				
1-Accurate	62	63	64	57	74	62	64	65	66	57	64	68	52	66	61	60	65	68	57	65
2a-Omisson of communion increase	9	6	6	8	10	6	7	7	11	5	8	7	9	5	8	17	8	7	7	5
2b-Omission of competence equality	7	7	3	6	3	7	11	7	6	5	8	7	4	8	11	0	8	7	10	7
3a-Overspecification to prescriptions	4	4	6	5	0	4	1	8	0	14	4	3	9	13	0	0	5	3	3	7
3b-Generalization	2	2	0	2	0	1	6	3	0	0	3	1	9	0	8	6	0	1	0	7
3c-Aspect not included in study	6	13	11	12	2	10	4	3	9	10	4	7	13	5	0	9	5	4	13	7
3d-Contradiction	10	5	11	10	10	11	6	6	9	10	10	8	4	3	11	9	11	11	10	2

Note. Gender included W = woman, M = man, and uk = unknown (d = diverse were excluded from this table given the small number [$n = 4$]). Seniority of author included S = student, PD = postdoc, AsP = assistant professor, AcP = associate professor, FP = full professor, PI = practice or industry, uk = unknown. Percentages in some columns may not total 100 due to rounding.

Citation Accuracy Depending on Discipline and Type of Citing Report

An exploratory analysis examined whether the citation accuracy differed by scientific discipline, including psychology (N=105 citations), management (N=33), medicine (N=12), social sciences (N=16), gender studies (N=12), computer science (N=6), sociology (N=6), and multidisciplinary (N=7). Disciplines with less than 5 citations were omitted from this analysis. As shown in Supplement C in the online supplement, results revealed the highest accuracy rate overall in gender studies (80%), sociology (70%), and psychology (66%) and the highest inaccuracy rates overall in computer science (33%), sociology (30%), and medicine (27%).

Finally, as shown in Supplement D in the online supplement, the overall association between degree of citation accuracy and the type of report (e.g., journal articles, dissertations, working papers, preprints) was nonsignificant, $\chi^2(16)=19.64$, $p=.240$.

Discussion

By providing a case study of the fragility of scientific knowledge of the phenomena of gender, this article reports the extent to which a recent and influential article on change in gender stereotypes over time has been miscited in the scientific literature. Our analysis of the 751 reports that cited the Eagly et al. (2020) findings by November 2023 showed that concerningly large percentages of citations distorted the original findings by either omitting important results, or sometimes contradicting the reported findings. Specifically, across all citations, 59% were accurate, 9% somewhat accurate, and a notably high rate of 32% were inaccurate, although the inaccuracy rate varied substantially across the three stereotype dimensions. Even though the descriptive nature of these findings disallows conclusions about the causes of this profusion of miscitations, comparisons to rates in other domains are helpful. Not only does the overall error rate revealed by our analysis exceed the 19% rate reported in the recent wide-ranging examination of miscitations in top psychology journals (Cobb et al., 2024), but also this rate exceeds the inaccuracy reported for some other fields, including 12–14.5% of citations being "seriously incorrect" in medicine (Jergas & Baethge, 2015; Mogull, 2017) and 8% being "entirely incorrect" in educational research (Lazonder & Janssen, 2022).

Our analyses revealed that inaccurate citations most often misrepresented the findings on agency. Specifically, 37% of the citing articles were inaccurate, with 21% of all agency citations directly contradicting the actual finding of continued belief in men's greater agency ever since the 1940s. The miscitations, for example, stated that the male agency stereotype had faded over time or that women and men are now seen as similarly agentic. Although this reported trend over time is not accurate according to the representative public opinion polls analyzed by Eagly et al. (2020), this belief aligns with lay theories people may have about how stereotypes should change when society becomes more gender equal, as we discuss below. For the communion citations, 25% ignored the actual finding that the female advantage has become even stronger over time. These citations, for example, although indicating that women are regarded as more communal than men, omitted the important nuance of the marked stereotypical increase over the years.

Finally, for the competence citations, 18% ignored that the most frequent current response was gender equality, that is, that women and men are viewed as similarly competent. These inaccurate citations, for example, stated that competence is now associated more with women than with men or that women are increasingly believed to be more competent than men, thus omitting the important nuance that belief in competence equality has increased over time and now represents most U.S. respondents. Illustrating this tendency is the example of the competence item "intelligence," for which belief in equal intelligence increased from only 35% in 1946, to 43% in 1995, and to an impressive 86% in 2018.

The reasons why the miscitation rates for gender stereotype findings exceeded those for some other research topics and fields remains an open question. One possibility is that people, including the authors of journal articles, tend to hold misleading beliefs about changes in gender over time. In evidence, past experimental research found that people believed women were becoming as agentic as men because of progress toward gender equality in role occupancies, whereas they also believed that women's greater communion was unchanging (see Diekman & Eagly, 2000). These lay theories may interfere with authors' understanding and remembering of findings from poll data showing how stereotypes have changed over time. Thus, authors' belief that women are as agentic as men might promote misremembering and misciting the demonstrated absence of change in the agency stereotype. Also, authors' belief that female advantage in communion is fixed and unchanging might promote their misremembering and misciting the demonstrated increase in women's stereotypical communion relative to that of men. Such errors are consistent with decades of research on cognitive distortions showing that personal beliefs, attitudes, and expectations can interfere with attention to information and bias its interpretation and recall (e.g., Chaiken et al., 1989; Eagly, 1992; Eagly et al., 2001).

Our finding that citation accuracy was unrelated to the gender of the first and last author refutes explanations of

citation inaccuracy arising from authors' gender. For example, women's greater engagement with gender issues might promote greater accuracy, or, alternatively, women and men might more accurately cite findings that are aligned with positive views about their own gender ingroup. The lack of such relations to author gender aligns with the overall similarity of gender stereotypes reported by women and men respondents (Eagly et al., 2020).

Accuracy was also unrelated to seniority of the first and last author, suggesting that greater scientific training and experience do not mitigate citation errors. Yet, the articles' findings were more accurately cited in disciplines that have long traditions researching gender such as gender studies and psychology than in disciplines not typically focusing on gender such as computer science and medicine.

Finally, citation accuracy was unrelated to journals' impact factor, which often serves as a metric of scientific quality, despite questioning of its validity (Chawla, 2018). Relatedly, inaccuracy did not differ between types of reports and was not, for example, lower in peer-reviewed journal articles compared to Bachelor theses. These results suggest that external circumstances such as publication outlet or academic experience were likely not key drivers of the citation errors.

More broadly, miscitation might also be fueled by pressure to publish prolifically to receive tenure or receive other positive outcomes in many contemporary universities. A high pressure to publish can lead researchers to take shortcuts, especially for those who do not have resources for reading their cited articles carefully or double checking their work (see Bauerlein et al., 2010; Simkin & Roychowdhury, 2006).

A final possibility is that the high rate of miscitation occurred because of the lead sentence in the Discussion of the Eagly et al. (2020) article, stating that "challenging traditional claims that stereotypes of women and men are fixed or rigid, our study joins others in finding stereotypes to be flexibly responsive to changes in group members' social roles" (Eagly et al., 2020, p. 310). An isolated reading of this statement, combined with the erroneous belief that women's and men's social roles have become similar, might lead some readers to wrongly infer that multiple dimensions of gender stereotyping—including agency—have changed; when the findings actually show stability in men's greater agency and an increase in women's communion. Yet, the abstract precisely described all of the study's main findings and thus should have served as a deterrent to inaccurate citation (Eagly et al., 2020, p. 301).

Limitations and Future Research Directions

Our analysis of the existing citations in the literature is not without limitations. First and foremost, this research does not provide definitive evidence for why the miscitation rates for gender stereotype findings exceeded those for some other research topics and fields. Therefore, we could only speculate about plausible reasons. Future research might develop samples of articles that challenge established assumptions or contradict established theories and findings in psychology. Investigating the mechanisms behind miscitation could provide valuable insights for training of researchers to reduce miscitations in scientific work.

Practice Implications

This case study, along with the survey by Cobb et al. (2024), conveys the fragility of at least a portion of scientific knowledge in psychology. This fragility has long been a theme in feminist criticism of science (see review by Eagly & Riger, 2014). Some feminists have thus emphasized the multiple biases in scientific research that call for triangulating findings across communities of psychologists who critique each other's work (e.g., Crawford & Kimmel, 1999). From a more radical postmodernist perspective, some feminists have argued that science reflects the social and political position of those who produce it and thus has a weak claim to valid representation of external reality (e.g., Hare-Mustin & Marecek, 1990). Whatever the merits of these broad claims, this case history on miscitation provides a detailed report on one of the many ways that science can go astray.

Conclusion

This case study provides a wake-up call concerning the corruption of scientific knowledge that can occur as research findings make their way through the preconceptions and prejudices of many of the researchers who cite them and carry their message into a body of scientific knowledge on gender. In fact, the notably high rate of miscitation that we have discovered in this case study is astonishing. Compared to the base rate of 19% erroneous citations of findings reported in top psychology journals (Cobb et al., 2024), our result of 32% inaccurate and 9% somewhat inaccurate citations of the Eagly et al. (2020) stereotype findings reveals a serious barrier to the orderly cumulation of findings. Miscitation was particularly severe for agency, the stereotype dimension for which the finding of no change in men's greater agency over time most strongly competes with people's lay theory that assumes a reduction over time.

In the current case, inaccurate assumptions about change in gender stereotypes apparently led to considerable citation inaccuracy, thus interfering with the accurate cumulation of knowledge. The lesson that emerges is that carelessness, likely amplified by pressure to publish a lot to promote

one's academic career, can compromise scientific writing, which requires careful attention to publications to prevent false assumptions from obscuring actual findings and contaminating the scientific knowledge base.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11199-025-01561-x.

Acknowledgements We thank Danielle Herzog, Isabelle Bartholomä, and Emily Läderach for their assistance with the data collection and coding of the citations.

Funding Open access funding provided by University of Bern
Open access funding provided by University of Bern. This work was supported, in part, by a postdoctoral fellowship grant from the Swiss National Science Foundation (No P400PS_199273) awarded to Christa Nater.

Data availability All data, analysis code, and research materials are available at OSF (https://osf.io/sfkp2). The preregistration of the coding procedure and analysis strategy are also available on OSF (https://osf.io/dy5tf).

Declarations

Compliance with Ethical Standards The manuscript adheres to ethical guidelines specified in the APA Code of Conduct. The research reported in this manuscript was exempt from institutional ethics board review.

Conflict of interest The authors declare that there are no potential conflicts of interest concerning the research, authorship, or publication of this article.

References

Bauerlein, M., Gad-el-Hak, M., Grody, W., McKelvey, B., & Trimble, S. W. (2010, June 13). *We must stop the avalanche of low-quality research.* The Chronicle of Higher Education. https://www.chronicle.com/article/we-must-stop-the-avalanche-of-low-quality-research/

Blau, F. D., & Winkler, A. E. (2018). Women, work, and family. In S. L. Averett, L. M. Argys, & S. D. Hoffman (Eds.), *The Oxford handbook of women and the economy* (pp. 395–424). Oxford University Press.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman, & J. A. Bargh (Eds.), *Unintended thought: Limits of awareness, intention, and control* (pp. 212–252). Guilford Press.

Chawla, D. S. (2018). *What's wrong with the journal impact factor in 5 graphs.* Nature Index. https://www.nature.com/nature-index/news/whats-wrong-with-the-jif-in-five-graphs

Cobb, C. L., Crumly, B., Montero-Zamora, P., Schwartz, S. J., & Martínez, C. R. Jr. (2024). The problem of miscitation in psychological science: Righting the ship. *American Psychologist*, *79*(2), 299–311. https://doi.org/10.1037/amp0001138

Crawford, M., & Kimmel, E. (1999). Promoting methodological diversity in feminist research. *Psychology of Women Quarterly*, *23*(1), 1–6. https://doi.org/10.1111/j.1471-6402.1999.tb00337.x

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*, 1171–1188. https://doi.org/10.1177/0146167200262001

Eagly, A. H. (1992). Uneven progress: Social psychology and the study of attitudes. *Journal of Personality and Social Psychology*, *63*(5), 693–710. https://doi.org/10.1037/0022-3514.63.5.693

Eagly, A. H., & Riger, S. (2014). Feminism and psychology: Critiques of methods and epistemology. *American Psychologist*, *69*(7), 685–702. https://doi.org/10.1037/a0037372

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology* (pp. 458–476). Sage.

Eagly, A. H., Kulesa, P., Chen, S., & Chaiken, S. (2001). Do attitudes affect memory? Tests of the congeniality hypothesis. *Current Directions in Psychological Science*, *10*(1), 5–9. https://doi.org/10.1111/1467-8721.00102

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*, 301–315. https://doi.org/10.1037/amp0000494

Hamblin, T. J. (2010, December 23). Spinach—I was right for the wrong reason. *Mutations of Mortality*. http://mutated-unmuated.blogspot.com/2010/12/spinach-i-was-right-for-wrong-reason.html

Hare-Mustin, R. T., & Marecek, J. (1990). Gender and the meaning of difference: Postmodernism and psychology. In R. T. Hare-Mustin, & J. Marecek (Eds.), *Making a difference: Psychology and the construction of gender* (pp. 22–64). Yale University Press.

Hinton, P. R. (2020). *Stereotypes and the construction of the social world.* Routledge.

Hoffower, H. (2019, July 18). *Women are now seen as just as competent as men, but less ambitious—And it's a good and bad thing.* Business Insider. https://www.businessinsider.com/women-versus-men-intelligence-competence-compassion-ambition-workplace-leadership-2019-7

Jergas, H., & Baethge, C. (2015). Quotation accuracy in medical journal articles—a systematic review and meta-analysis. *PeerJ*, *3*, Article e1364. https://doi.org/10.7717/peerj.1364

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*, 371–392. https://doi.org/10.1037/a0037215

Lazonder, A. W., & Janssen, N. (2022). Quotation accuracy in educational research articles. *Educational Research Review*, *35*, Article 100430. https://doi.org/10.1016/j.edurev.2021.100430

Mogull, S. A. (2017). Accuracy of cited facts in medical research articles: A review of study methodology and recalculation of quotation error rate. *Plos One*, *12*(9). https://doi.org/10.1371/journal.pone.0184727. Article e0184727.

Oxford English Dictionary. (n.d.). Stereotype. Retrieved September 25, 2024, from https://www.oed.com/dictionary/stereotype_n?tab=factsheet#20628569

Oyserman, D., & Yan, V. X. (2019). Making meaning: A culture-as-situated-cognition approach to the consequences of cultural fluency

and disfluency. In D. Cohen, & S. Kitayama (Eds.), *Handbook of cultural psychology* (2nd ed., pp. 536–565). Guilford Press.

Rekdal, O. B. (2014). Academic urban legends. *Social Studies of Science*, *44*, 638–654. https://doi.org/10.1177/0306312714535679

Salam, M. (2019, July 30). Americans finally consider women as competent as men. *The New York Times*. https://www.nytimes.com/2019/07/30/us/women-men-competence.html

Schmader, T., & Nater C. (in press). Gender. In D. Gilbert, S. Fiske, E. Finkel, & W. B. Mendes (Eds.), *The handbook of social psychology* (pp. 1–120). Princeton University Library.

Simkin, M., & Roychowdhury, V. (2006). Do you sincerely want to be cited? Or: Read before you cite. *Significance*, *3*(4), 179–181. https://doi.org/10.1111/j.1740-9713.2006.00202.x