

The gender balance of expert sources quoted by UK news outlets online

An analysis by the Centre for the Study of Media,
Communication and Power, on behalf of the
Global Institute for Women's Leadership

Findings

This analysis was conducted by the [Centre for the Study of Media, Communication and Power](#) at King's College London, on behalf of the Global Institute for Women's Leadership. It looks at the gender balance of people quoted as experts in online news articles by the main UK news outlets, across eight categories of coverage. Overall, **77% of these expert sources are men, while 23% are women.** The analysis was conducted using a representative sample of articles over a seven-day period in April 2018.

The table below contains a more detailed breakdown of the gender balance of expert sources according to type of news coverage.

Category			Source gender		Total
			Male	Female	
Social Policy	Count		35	32	67
	% within Article Type/Focus		52.2%	47.8%	100.0%
Foreign	Count		229	65	294
	% within Article Type/Focus		77.9%	22.1%	100.0%
Foreign Politics	Count		375	57	432
	% within Article Type/Focus		86.8%	13.2%	100.0%
Politics	Count		315	141	456
	% within Article Type/Focus		69.1%	30.9%	100.0%
Business/ Finance	Count		268	45	313
	% within Article Type/Focus		85.6%	14.4%	100.0%
Science/ Health	Count		110	46	156
	% within Article Type/Focus		70.5%	29.5%	100.0%
Tech	Count		47	15	62
	% within Article Type/Focus		75.8%	24.2%	100.0%
Nature/ Environment	Count		33	12	45
	% within Article Type/Focus		73.3%	26.7%	100.0%
Total	Count		1412	413	1825
	% within Article Type/Focus		77.4%	22.6%	100%

Methodology

Key numbers - sample

- Sample period: 19th – 25th April 2018
- Number of articles in present analysis: 1,112
- Number of expert sources in these articles: 1,825
- Total eligible news articles scraped from the websites of the main UK broadcast and print news media between 19th and 25th April 2018: approximately 22,500

Data collection

All content was collected using the Steno computational analysis tool, which uses pre-programmed scrapers to collect and store all URLs published on selected news sites over a defined time period. For this project, the period of analysis was a one-week sample from Thursday 19th April to Wednesday 25th April 2018. The full dataset from which this analysis is derived therefore consists of all unique articles published on selected UK news sites during a one-week period.

The selected publications from which URLs were scraped cover the main television public service broadcasters and the sites of UK national newspapers. The list of publications is as follows:

BBC (bbc.co.uk/news)	Broadcast
Channel 4 News (channel4.com/news)	Broadcast
ITV (itv.com/news)	Broadcast
Sky News (news.sky.com)	Broadcast
Daily Mail & Mail on Sunday (dailymail.co.uk)	Legacy Print
Daily Star & Daily Star Sunday (dailystar.co.uk)	Legacy Print
Daily Express & Sunday Express (express.co.uk)	Legacy Print
Financial Times (ft.com)	Legacy Print
Guardian (theguardian.com)	Legacy Print
Independent (independent.co.uk)	Legacy Print
The i (inews.co.uk)	Legacy Print
Daily Mirror, Sunday Mirror and Sunday People (mirror.co.uk)	Legacy Print
Sun & Sun on Sunday (thesun.co.uk)	Legacy Print
Daily Telegraph & Sunday Telegraph (telegraph.co.uk)	Legacy Print
The Times & The Sunday Times (times.co.uk)	Legacy Print

Once all data was collected for the week-long time period, certain content was filtered out and removed from the dataset. This includes: sports coverage (since certain outlets, such as the BBC and Sky News, host this content on different domains); reviews (product and entertainment, including television programme summaries); and non-news content, such as information pages (e.g. TV schedules), crosswords, competitions, recipes and cookery, gardening, etc. Audio/video-only articles, including podcasts, were not included in the analysis, which is of text content in online news articles. Where articles contained videos alongside multiple paragraphs of text, the article was included in the analysis on the basis of the text.

In total, approximately 22,500 articles were collected in this full dataset after sports and ineligible articles were removed (though a small number which had been missed by filters were later removed from the sample – see below).

Sampling and analysis

For the analysis, a 12.5% sample of articles from each publication was randomly selected. This sample was then scanned to identify any ineligible articles that had been missed by the filtering process described above. A number of sports articles were identified and removed (mostly from *Daily Mail* wire-copy articles), as well as a small number of additional ineligible articles.

The sample for content analysis consisted of 2,671 articles. Of these, 408 contained no quoted sources; these were removed, resulting in a total sample of 2,263 articles. Articles were later allocated to certain categories on the basis of the central focus of the story (see table below). The number of articles used in the current analysis is 1,112, in which 1,825 sources are featured.

The gender of sources was recorded on the basis of pronouns and names within article texts. Where the gender of a source was not clear, individuals were identified through their host institutions. Anonymous sources with identifiable gender ('spokesman', 'spokeswoman', 'he said', 'she said,') were recorded as such. Anonymous sources with no identifiable gender ('spokesperson'; 'source') were not recorded; group statements were also not recorded). Journalists were not recorded as sources unless they were quoted from other publications (e.g. *New York Times* journalists quoted as sources on American politics in *The Times*), or if they were published in comment pieces.

Sources were identified as 'expert' on the basis of whether they were identified via their job title or institutional affiliation or by other signals of authority. Signifiers of 'expert' status were defined as:

- Job title (e.g. CEO, Chair, analyst, agent, academic title)
- Institutional affiliation (spokesperson for public/private/charitable organisation)
- Political title (e.g. MP in domestic news; government or opposition politicians in foreign news)

Articles were coded according to topic (see table below).

The unit of analysis is quoted sources. Within each article, every unique identifiable source was recorded once. For example, if in one article source A was quoted multiple times while source B was quoted once, these would both be recorded equally. The analysis therefore records *which sources feature in a given article*.

The variables recorded for each quoted source in the analysis were as follows:

Variable	Description and Values
'Publication'	Publication in which the article appears (e.g. BBC, Channel 4, Express)
'Headline'	Headline of article
'Comment'	Whether or not an article is a comment piece or factual reporting. Comment pieces were recorded as featuring a single, expert, source.
'Article Focus'	<p>Main subject of the article in which the source appears, based on a pre-set list of categories:</p> <ul style="list-style-type: none"> • 'Politics' – articles about UK domestic UK politics where elected officials are represented • 'Social policy' – articles about specific UK social policy areas (health, welfare and education) where elected officials do not feature • 'Business/Finance' – articles about domestic or international business, finance, and/or economics • 'Foreign politics' – articles about politics, policy or conflict internationally, where political figures feature • 'Foreign' – foreign news with no policy dimension, includes foreign crime stories, but not 'entertainment/celebrity' • 'Science/Health' – articles about science or scientific research, including medical research (not including climate or environmental research) • 'Tech' – articles about technology or engineering, including social media/tech industries • 'Nature/Environment' – articles about the natural world, the environment and climate change
'Source Gender'	Identified gender of source (see above for methods of identification)
'Expert/Non-Expert'	Whether source is identified as 'expert' or 'non-expert' (see above for criteria)

The remainder of the articles in the sample were allocated categories not analysed in the present study. These were: ‘Crime’, ‘Human interest’, ‘Entertainment/celebrity’, ‘Accident/Tragedy’, ‘Lifestyle’, ‘Tabloid/Bizarre,’ and a catch-all ‘Other’ category for remaining stories.

Caveats

The analysis consists of a list of *who gets to speak as an expert source* in the online output of the UK’s largest media outlets. No distinction is drawn between length of quote and prominence within articles – it is a list of which identifiable sources feature in the sampled articles. The prevalence of sources whose gender is unidentifiable is relevant, but outside the scope of this analysis.

‘Politics’ articles were categorised differently from ‘Social policy’ articles in terms of whether party political or elected figures were included. ‘Social policy’ articles are inherently political, but a distinction was drawn between those articles that contain governmental or party-political activity, and those covering areas of social policy that do not.

The analysis contains no judgement or analysis of why there may be gender imbalances among sources or in the use of ‘expert’ and ‘non-expert’ sources in the publications sampled here – it simply records the gender and status of those identifiable sources who do speak. The gender balance in certain industries (e.g. tech and engineering, finance) is a key factor here which lies outside the control of media outlets. However, in the context of *the expert sources that make up the news articles consumed by audiences* it is valid to analyse the voices and sources that audiences are exposed to.

Acknowledgements

The Global Institute for Women's Leadership would like to thank Dr Gordon Ramsay, Deputy Director of Centre for the Study of Media, Communication and Power, for carrying out this analysis

giwl.kcl.ac.uk
@GIWLkings
#womensleadership
giwl@kcl.ac.uk