Supplementary Information for

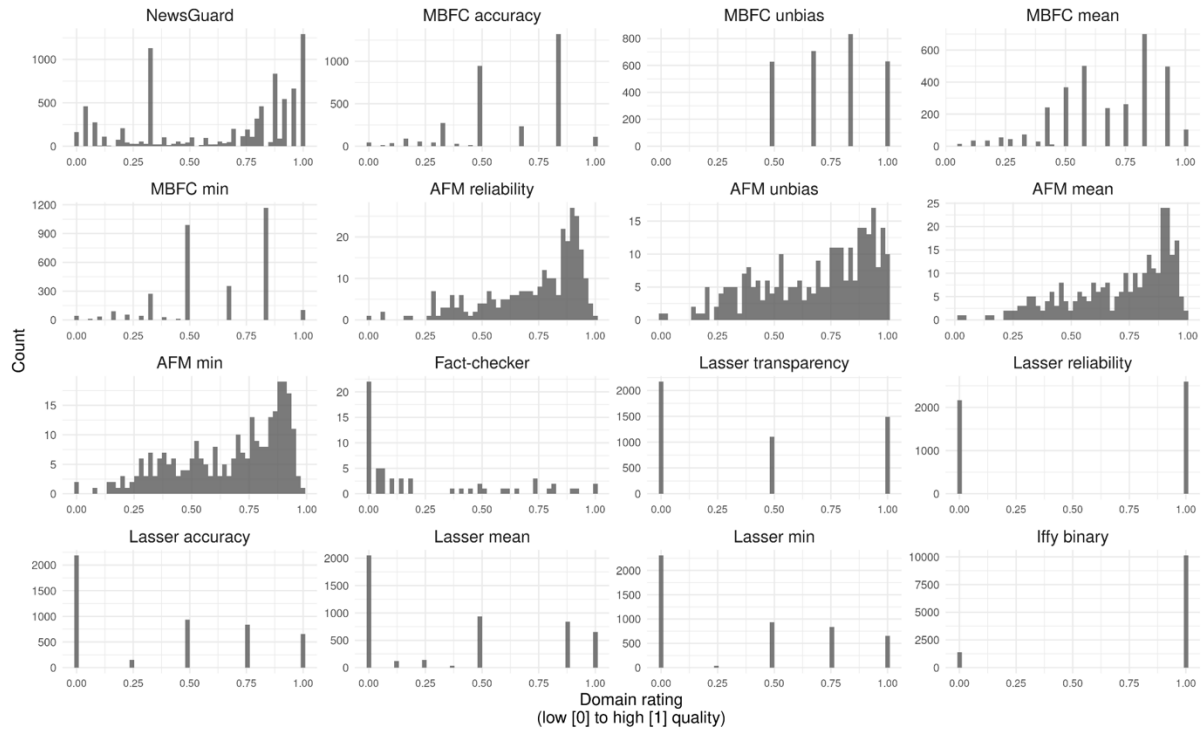## High level of correspondence across different news domain quality rating sets

Figure S1. Distributions of ratings for each source. Each panel is one source. The quality ratings on the x-axis range from 0 (lowest quality) to 1 (highest quality). MBFC: Media Bias/Fact Check. AFM: Ad Fontes Media. Lasser: (1). Iffy: Iffy index of unreliable sources.
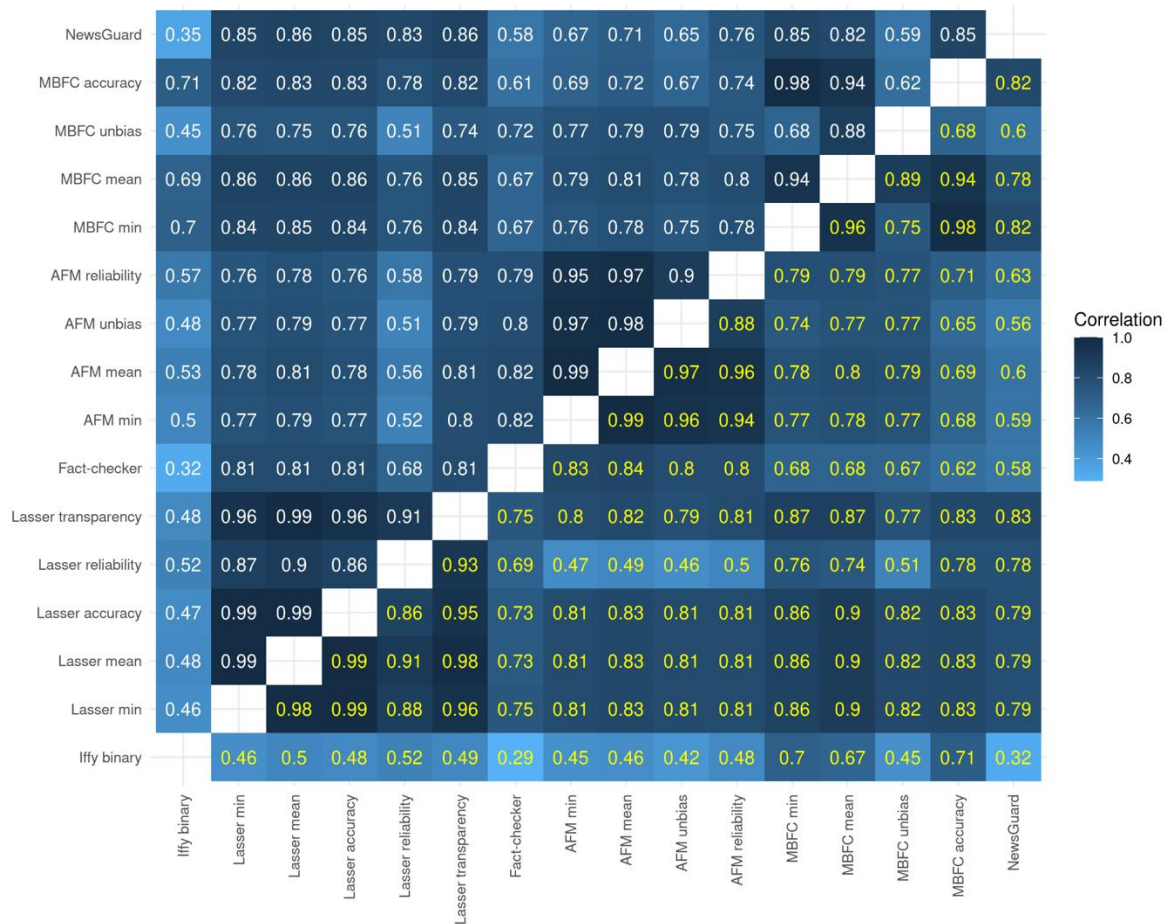
Figure S2. Domain rating correlations. Values in the upper triangle are Pearson's *r* and values in the lower triangle are Spearman's *ρ*. MBFC: Media Bias/Fact Check. AFM: Ad Fontes Media. Lasser: (1). Iffy: Iffy index of unreliable sources.
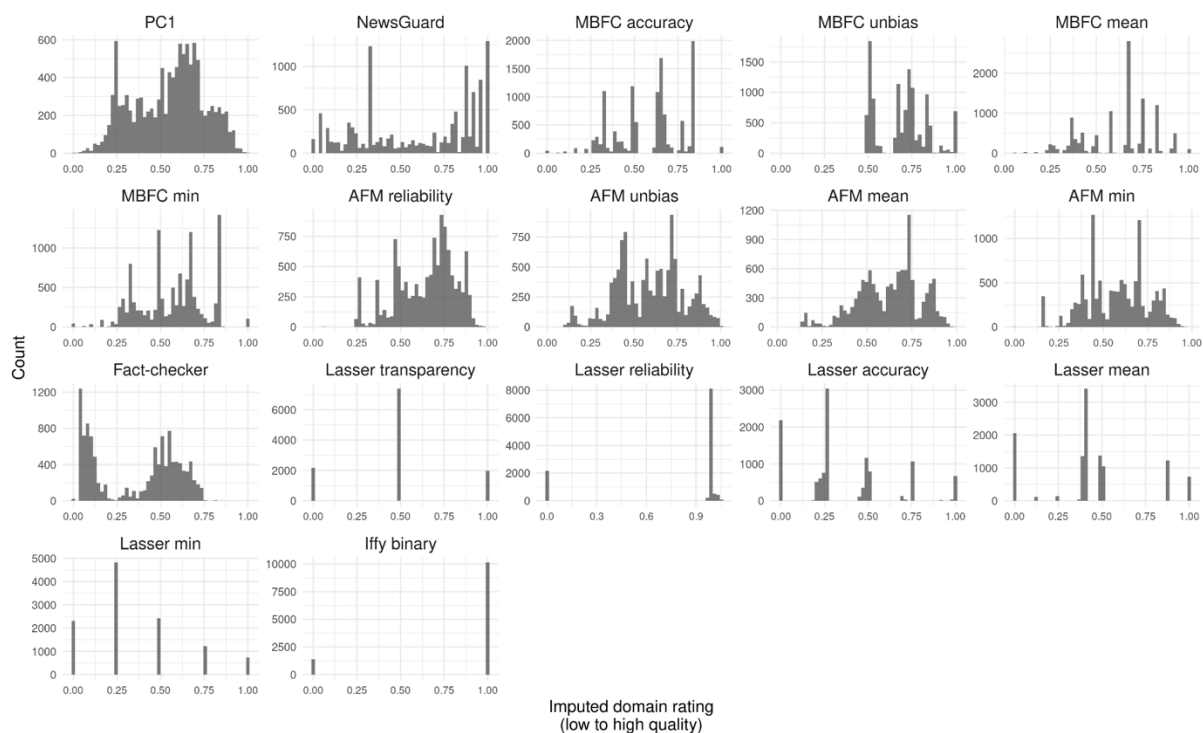
Figure S3. Distributions of imputed ratings for each source. PC1: first principal component. MBFC: Media Bias/Fact Check. AFM: Ad Fontes Media. Lasser: (1). Iffy: Iffy index of unreliable sources.
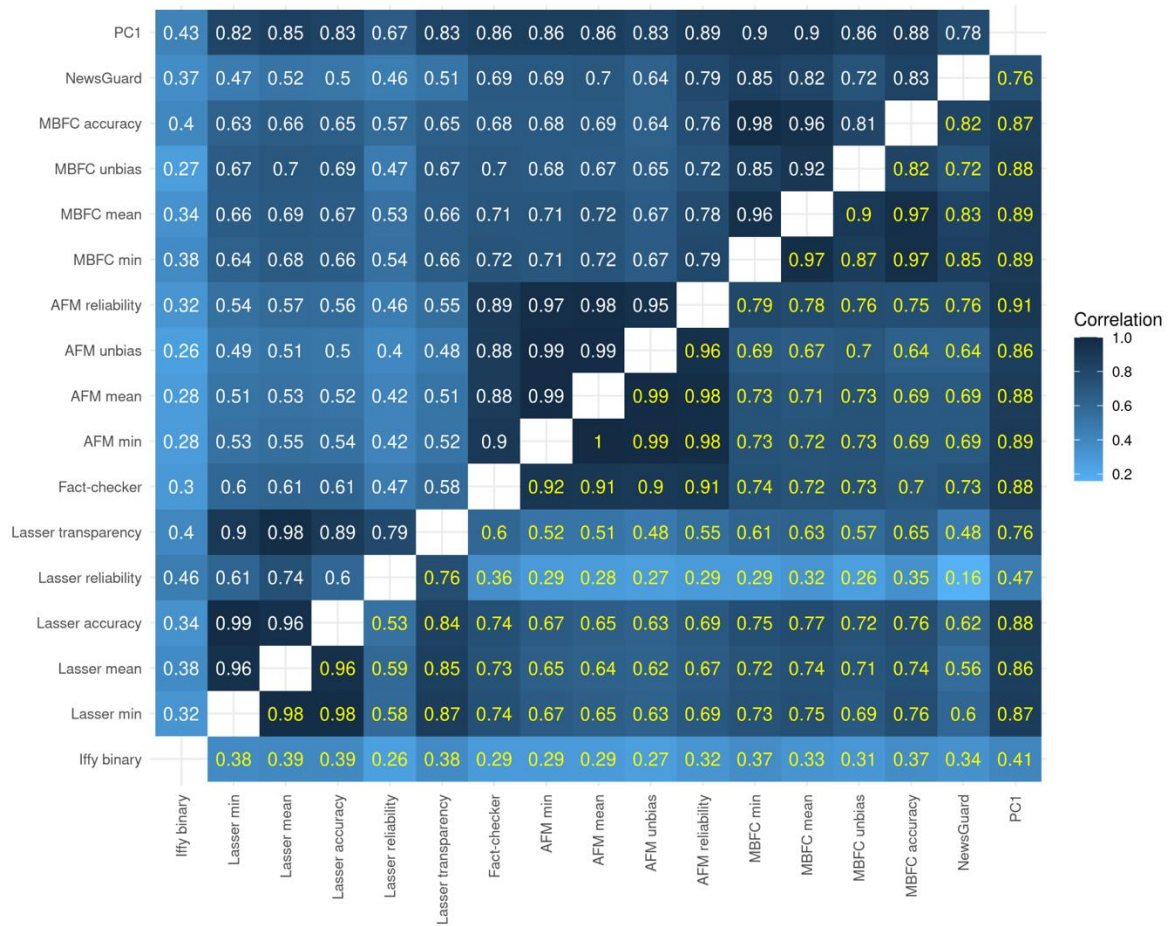
	Iffy binary	Lasser min	Lasser mean	Lasser accuracy	Lasser reliability	Lasser transparency	Fact-checker	AFM min	AFM mean	AFM unbias	AFM reliability	MBFC min	MBFC mean	MBFC unbias	MBFC accuracy	NewsGuard	PC1
PC1	0.43	0.82	0.85	0.83	0.67	0.83	0.86	0.86	0.86	0.83	0.89	0.9	0.9	0.86	0.88	0.78	
NewsGuard	0.37	0.47	0.52	0.5	0.46	0.51	0.69	0.69	0.7	0.64	0.79	0.85	0.82	0.72	0.83		0.76
MBFC accuracy	0.4	0.63	0.66	0.65	0.57	0.65	0.68	0.68	0.69	0.64	0.76	0.98	0.96	0.81		0.82	0.87
MBFC unbias	0.27	0.67	0.7	0.69	0.47	0.67	0.7	0.68	0.67	0.65	0.72	0.85	0.92		0.82	0.72	0.88
MBFC mean	0.34	0.66	0.69	0.67	0.53	0.66	0.71	0.71	0.72	0.67	0.78	0.96		0.9	0.97	0.83	0.89
MBFC min	0.38	0.64	0.68	0.66	0.54	0.66	0.72	0.71	0.72	0.67	0.79		0.97	0.87	0.97	0.85	0.89
AFM reliability	0.32	0.54	0.57	0.56	0.46	0.55	0.89	0.97	0.98	0.95		0.79	0.78	0.76	0.75	0.76	0.91
AFM unbias	0.26	0.49	0.51	0.5	0.4	0.48	0.88	0.99	0.99		0.96	0.69	0.67	0.7	0.64	0.64	0.86
AFM mean	0.28	0.51	0.53	0.52	0.42	0.51	0.88	0.99		0.99	0.98	0.73	0.71	0.73	0.69	0.69	0.88
AFM min	0.28	0.53	0.55	0.54	0.42	0.52	0.9		1	0.99	0.98	0.73	0.72	0.73	0.69	0.69	0.89
Fact-checker	0.3	0.6	0.61	0.61	0.47	0.58		0.92	0.91	0.9	0.91	0.74	0.72	0.73	0.7	0.73	0.88
Lasser transparency	0.4	0.9	0.98	0.89	0.79		0.6	0.52	0.51	0.48	0.55	0.61	0.63	0.57	0.65	0.48	0.76
Lasser reliability	0.46	0.61	0.74	0.6		0.76	0.36	0.29	0.28	0.27	0.29	0.29	0.32	0.26	0.35	0.16	0.47
Lasser accuracy	0.34	0.99	0.96		0.53	0.84	0.74	0.67	0.65	0.63	0.69	0.75	0.77	0.72	0.76	0.62	0.88
Lasser mean	0.38	0.96		0.96	0.59	0.85	0.73	0.65	0.64	0.62	0.67	0.72	0.74	0.71	0.74	0.56	0.86
Lasser min	0.32		0.98	0.98	0.58	0.87	0.74	0.67	0.65	0.63	0.69	0.73	0.75	0.69	0.76	0.6	0.87
Iffy binary		0.38	0.39	0.39	0.26	0.38	0.29	0.29	0.29	0.27	0.32	0.37	0.33	0.31	0.37	0.34	0.41

Figure S4. Imputed domain rating correlations. Values in the upper triangle are Pearson's *r* and values in the lower triangle are Spearman's *ρ*. PC1: first principal component. MBFC: Media Bias/Fact Check. AFM: Ad Fontes Media. Lasser: (1). Iffy: Iffy index of unreliable sources.
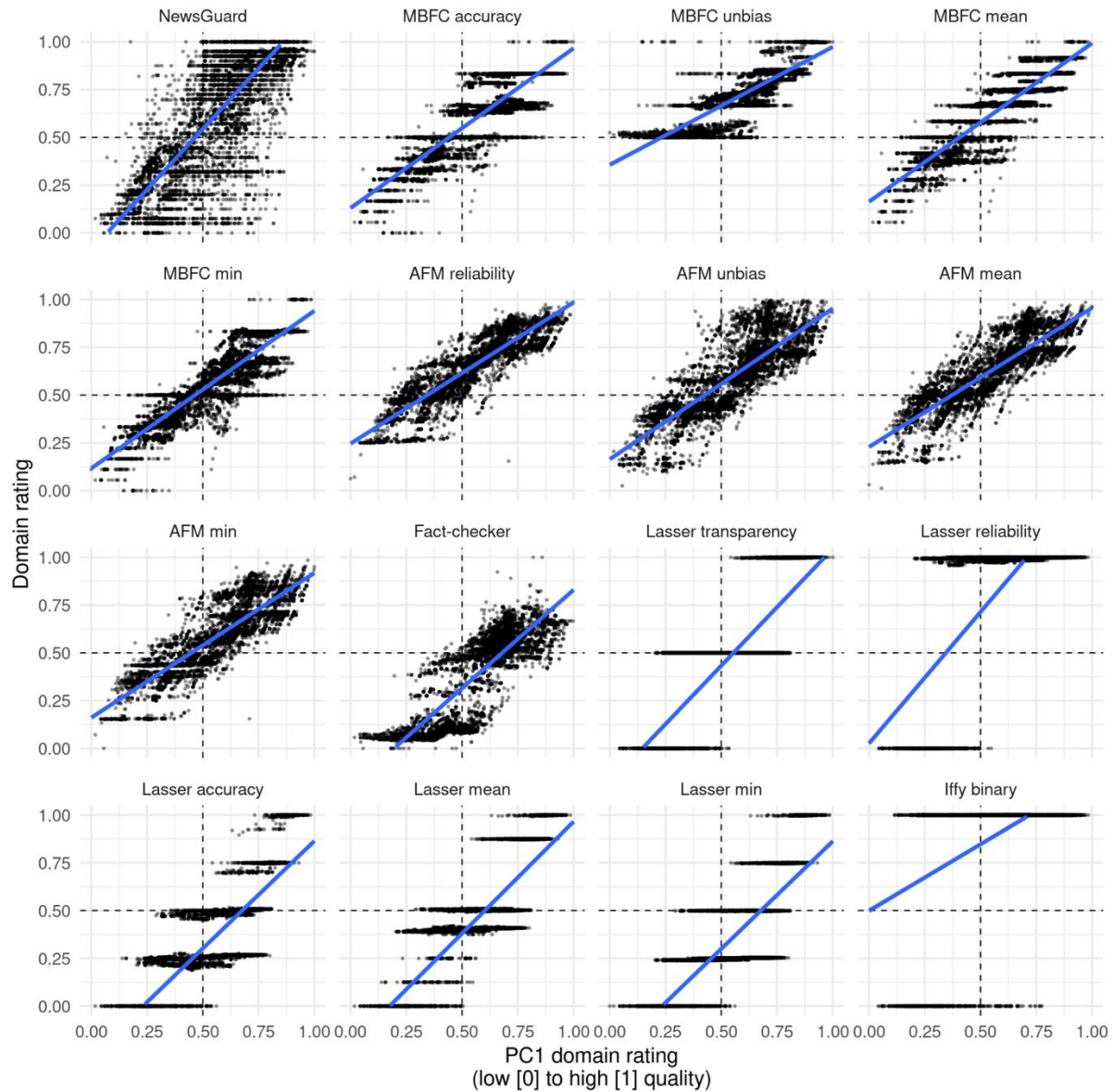
Figure S5. Scatterplots showing the relationships between the first principal component (PC1) and other imputed source ratings. Each dot is one domain.PC1: first principal component. MBFC: Media Bias/Fact Check. AFM: Ad Fontes Media. Lasser: (1). Iffy: Iffy index of unreliable sources.
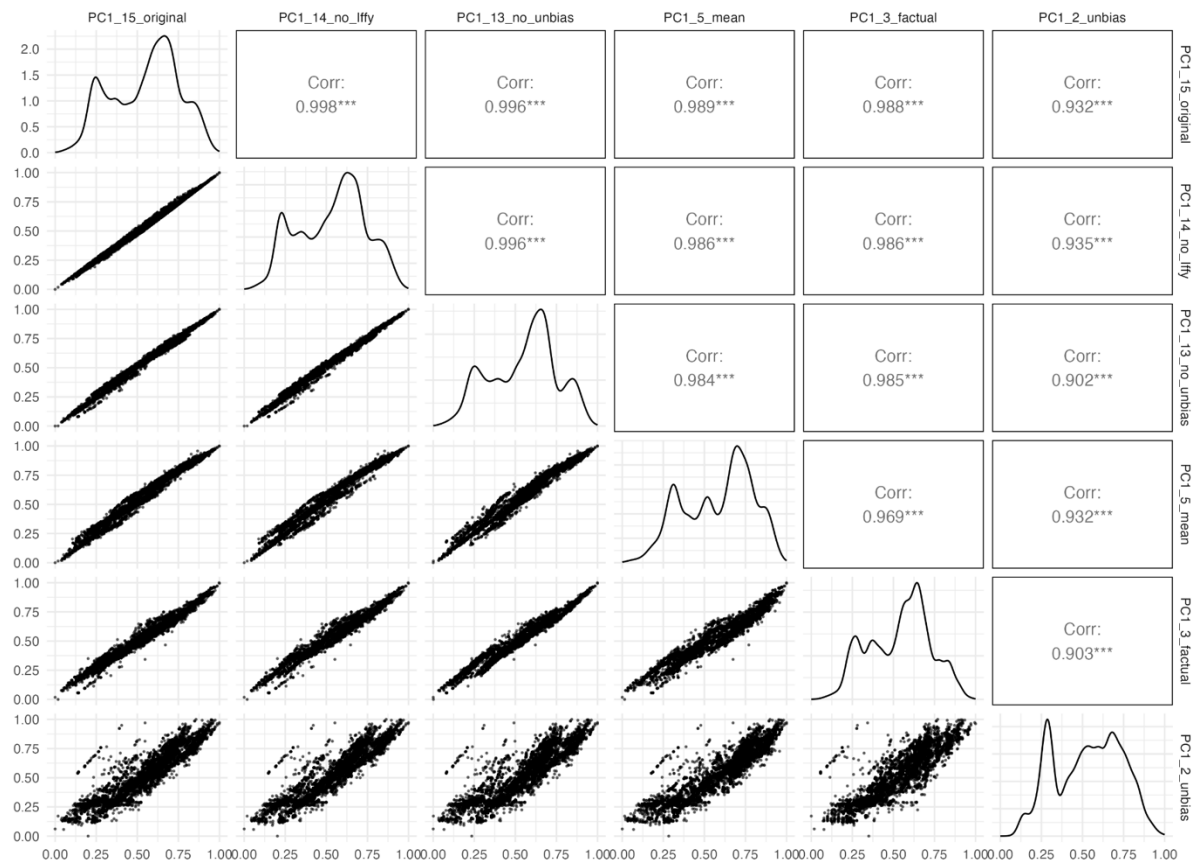
Figure S6. PC1 correlations for six PCAs that used different input vectors. The first column and row (PC1_15_original) is the PC1 vector reported in the main text. PC1_14_no_Iffy has 14 input vectors (excluding Iffy binary); PC1_13_no_unbias has 13 input vectors (excluding AFM unbias and MBFC unbias); PC1_5_mean has 5 input vectors (Fact-checker, AFM mean, MBFC mean, Lasser mean, Iffy binary); PC1_3_factual has 3 input vectors (AFM reliability, MBFC accuracy, Lasser accuracy); PC1_2_unbias has 2 input vectors (AFM unbias, MBFC unbias).
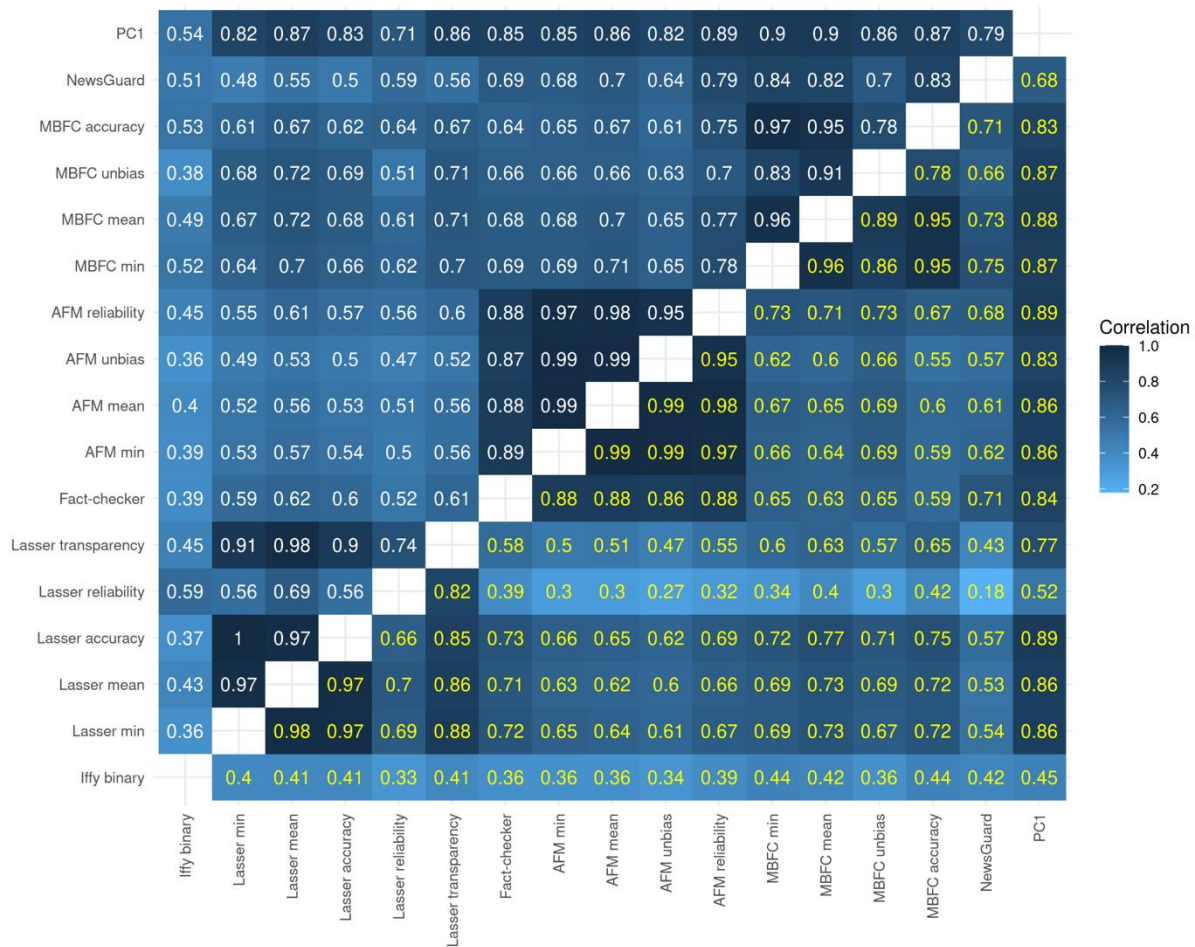
Figure S7. Imputed domain rating weighted correlations. Correlations are weighted by Open PageRank's domain importance metric. Values in the upper triangle are Pearson's *r* and values in the lower triangle are Spearman's *ρ*. PC1: first principal component. MBFC: Media Bias/Fact Check. AFM: Ad Fontes Media. Lasser: (1). Iffy: Iffy index of unreliable sources.

To validate the imputation method, we simulate data by drawing samples from a multivariate normal distribution to create a simulated dataset that has the same correlations as our imputed data (the number of samples is the number of domains and the covariance of this multivariate distribution mirrors the observed imputed domain correlations shown in Figure 5).

First, we perform PCA on this simulated dataset (without any missing values) and find that PC1 explains 67.95% of the variance in the data. Next, we introduce missing values such that the total number of missing values in each variable/source was identical to the original data, but were *missing randomly*. We perform imputation and PCA on this data, and find that PC1 explains 67.84% of the variance. Finally, we introduce missing values in the simulated dataset to recreate the non-random missing value structure in the original dataset. We then perform imputation and PCA, and find that PC1 explains 63.32% of the variance. Supplementary Figure 8 shows the two PC1s from the two imputed datasets correlated highly with the PC1 from the simulated data without missing values (left and middle panels). The PC1 from the dataset with data missing non-randomly (middle panel) corresponds better with the PC1 from the data without missing values ($r = 0.73$ and the correlation/regression line is closer to the diagonal), relative to PC1 from the dataset with data missing at random (left panel; $r = 0.64$). These simulations indicate that the imputation method works relatively well for our missing value data structure.
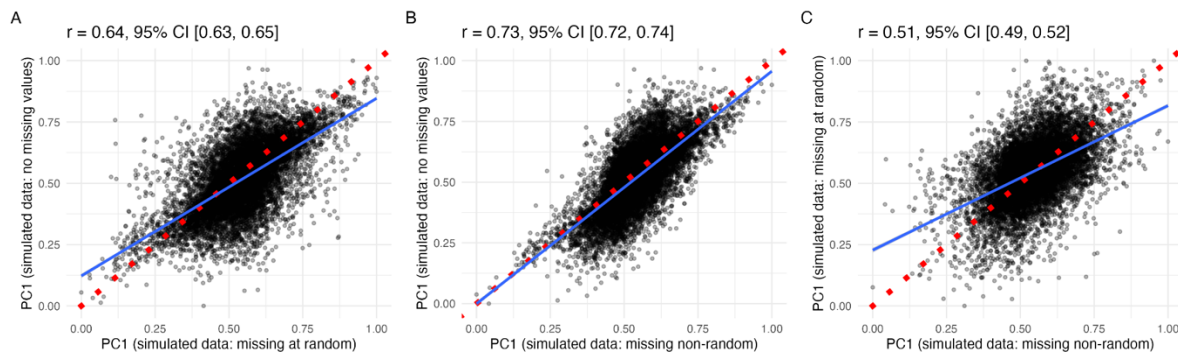

Figure S8. Correlation between PC1 from the simulated data without missing values and PC1 from the simulated data with data missing at random (**A**). Correlation between PC1 from the simulated data without missing values and PC1 from the simulated data with data missing non-randomly that replicates the missing data structure of the observed data (**B**). Correlation between PC1 from the simulated data with data missing at random and PC1 from the simulated data with data missing non-randomly (**C**). The blue line is the correlation/regression line; the red dotted line is the diagonal.

References

1. Lasser J *et al.* 2022. Social media sharing by political elites: An asymmetric American exceptionalism. *PNAS Nexus*