Supplementary Materials

Matthew Brown and Harsh Gupta

July 15, 2023

A Materials and Methods

A.1 Media outlet slant

We identify media outlets from Altmetrics (discussed below) and (Bakshy *et al.* (1)). (Bakshy *et al.* (1)) provides our measure of outlet slant, which we refer to as the BMA score. The BMA score captures the self-reported ideology of an outlet's online consumers. Each news article in their data receives an alignment score between -2 and 2, based on the average self-reported ideology of the users sharing the article on Facebook. Articles shared mostly by liberals receive negative scores, while those shared mostly by conservatives receive positive scores. The outlet's BMA score is the mean ideology of all its articles.

A.2 Citations of research by media outlets

We use Altmetric to construct our media citation data. Altmetric is a bibliometric database that scrapes the websites of thousands of major news outlets every day, counting every instance that a scientific paper is mentioned. Altmetric began scraping news sites in early 2013, and we conducted our search in December 2021. These dates are the bounds for our news sample. Over this time period, Altmetric's text mining capabilities improved, so we observe more citations in later years.

Most citations identified by the Altmetric scraping tool are from news articles that include hyperlinks to scientific journal websites papers or DOIs. A minority of citations are also identified via text mining. For a citation to be identified via text mining, the news article must include the scientist's name, a publication date, and the name of the publishing journal. Some prominent news outlets have few observations or are missing from the Altmetric database (for example, the Wall Street Journal). The reason is that Altmetric cannot scrape websites with restricted content such as paywalls.[1]

Many news organizations publish stories written by journalists at independent organizations, such as the Associated Press (AP). An article from an AP reporter about a topic of national interest might appear in dozens of local newspapers. Since individual news outlets have relatively little agency over the content that appears in these syndicated articles, we restrict our sample to articles that are only published in a single news outlet. In particular, we drop all observations for which multiple news articles share an identical title.

We *a priori* exclude two websites[2] that specialize in science coverage. We also drop news outlets with fewer that cover less than 20 scientific articles. Our final sample consists of 139 media outlets.

A.3 Scientific papers

Our primary data source for scientific articles is Dimensions ((2)). Dimensions is a large database that aggregates information on different features of scientific articles. This includes bibliometric information, funding data, patents, and so on.

For each day from 2015 to 2020, we queried the Dimensions for the thousand scientific articles with the most academic citations that were released on that day. Our initial sample contains $2,184,628$ scientific articles.

For each paper, we collect data on authors: names, university affiliations, fields, and locations. We also collect paper-level variables such as field of research, journal of publication, and year of publication. Last, we collect information on different quality measures of research: number of scientific citations, journal impact, research funding, patent citations, and policy citations.

A.3.1 Interpretation of quality measures

In our analysis, we use the aforementioned measures to capture the scientific quality of an article. In general, it is challenging to define and measure the quality of science. A key reason is that the quality of a scientific article is multifaceted. For example, scientific quality can be defined based on impact, plausibility, novelty, practical applicability, or relevance. Any measure of quality will likely miss important notions of quality. This

[1]Source: Personal correspondence with Altmetric
[2]globalresearch.ca and iflscience.com

incomplete measurement of scientific quality is important for our analysis, especially if we systematically exclude measures of quality that are valued by outlets of a certain ideology.

We attempt to overcome this limitation by developing a quality index that captures different notions of quality. Our index is based on number of scientific citations, quality of the journal, research funding, and patent citations for a scientific article. These measures capture quality from different perspectives. Scientific citations capture relevance from the perspective of other scientists. Research funding captures potential value from a funder's perspective. Patent citations capture the practical applicability of a scientific article. We contend that these measures capture a wide range of quality criteria. However, we recognize that our measure may still remain incomplete.

We construct our quality index by performing a Principal Component Analysis on the four measures discussed above. Our index is the first principal component.

A.4 Donation-based ideology scores

We use the Database on Money, Ideology, and Elections (DIME) (3) to construct ideology scores for scientific papers. DIME compiles public data on political contributions to U.S. politicians at the federal, state, and local levels from 1976-2014. Each contributor's donation behavior is summarized by a *Campaign Finance (CF) Score*, which places the contributor on a one-dimensional left-right scale (4).

A.5 Linking DIME and dimensions

In this section, we explain our procedure to link CF Scores to scientists, and then to papers.

We begin by dropping all DIME contributors whose most recent donation was earlier than 2006. Excluding these observations limits noise, since people who donated in the distant past are unlikely to be writing scientific papers in the time period we study. Next, we construct subsamples of recent contributors whose occupations indicate that they are likely to be researchers. In our *strict donor sample*, we include only those individuals whose self-reported occupation is "professor," "scientist," or "researcher." The strict donor sample contains 71,473 observations. In our *expanded donor sample*, we also include those whose occupation is "student," "physician," and "doctor." The expanded donor sample contains 248,778 observations.

Similarly, beginning from the Dimensions sample of authors, we construct a sample of unique authors. We restrict the sample to researchers who are affiliated with an institution in the U.S. This *U.S. author subsample* contains 4,492,819 observations.

We extract first names and last names from each donor and author. We then computed Levenshtein distances between author names and donor names. The algorithm for our preferred linking procedure is as follows.

- Using the expanded donor sample, consider all pairs of scientists and donors where the name distance is less than or equal to one.
- Select links where the scientist matches to a single donor.
- Select links where the donor is a part of the strict donor sample, and the donor and the author are in the same state.
- Select links where the donor is a part of the strict donor sample, and the donor's name exactly matches the author's name (distance is 0).
- Drop all unmatched scientists.

We are able to assign a CF score to $259,826$ scientists (5.7% of the U.S. author subsample) using our procedure.

Finally, we transform our individual-level ideology scores to paper-level scores. We define a paper's CF Score to be the average CF Score of the papers' authors for whom a link existed in DIME. The final linked sample consists of $661,923$ scientific papers with CF Scores.

A.5.1 Validation

Assuming the propensity to donate to scientists is close to that of the general U.S. population, the fraction of scientists with a match in the donor pool should roughly equal the fraction of Americans who are donors. The U.S. adult population in 2010 was about 235 million, and DIME contains records for 10.4 million unique contributors who made donations between 2006 and 2014, so about 4.4% of people appear in DIME.[3] We link 5.7% of our authors, which is commensurate with this fraction.

We conducted a manual audit to more rigorously evaluate our linking procedure. We randomly sampled 319 scientists from the dimensions sample, oversampling successful links.

[3]Source: ff

Within the sample of links, we also oversampled cases where there was a unique donor for the scientists since our procedure focuses on such links and we wanted to increase our ability to distinguish between candidate algorithms.

For each scientist, we conducted manual searches of the web and the DIME dataset to determine any true donor matches. We use these manually linked records as the ground truth against which we evaluate our linking procedure.

We used a receiver operating characteristic (ROC) curve to evaluate various procedures. The ROC curve plots the false positive rate (the fraction of truly negative observations that are coded as positive by the procedure) against the true positive rate (the fraction of truly positive observations that are coded as positive by the procedure). A perfect procedure would be at the top left corner, while a completely uninformative procedure would lie somewhere in the bottom right corner. If a procedure randomly matched scientists to donors, it would lie on the 45-degree line. Figure S1 presents such a ROC curve for several candidate algorithms, with the algorithm that we chose highlighted in blue.

We face a tradeoff between minimizing false positives and maximising true positives, so there were several procedures on the frontier of the ROC curve. We chose our preferred procedure because it presents a reasonable balance between the two objectives. In section E.2, we show that our findings are robust to alternative linking procedures.

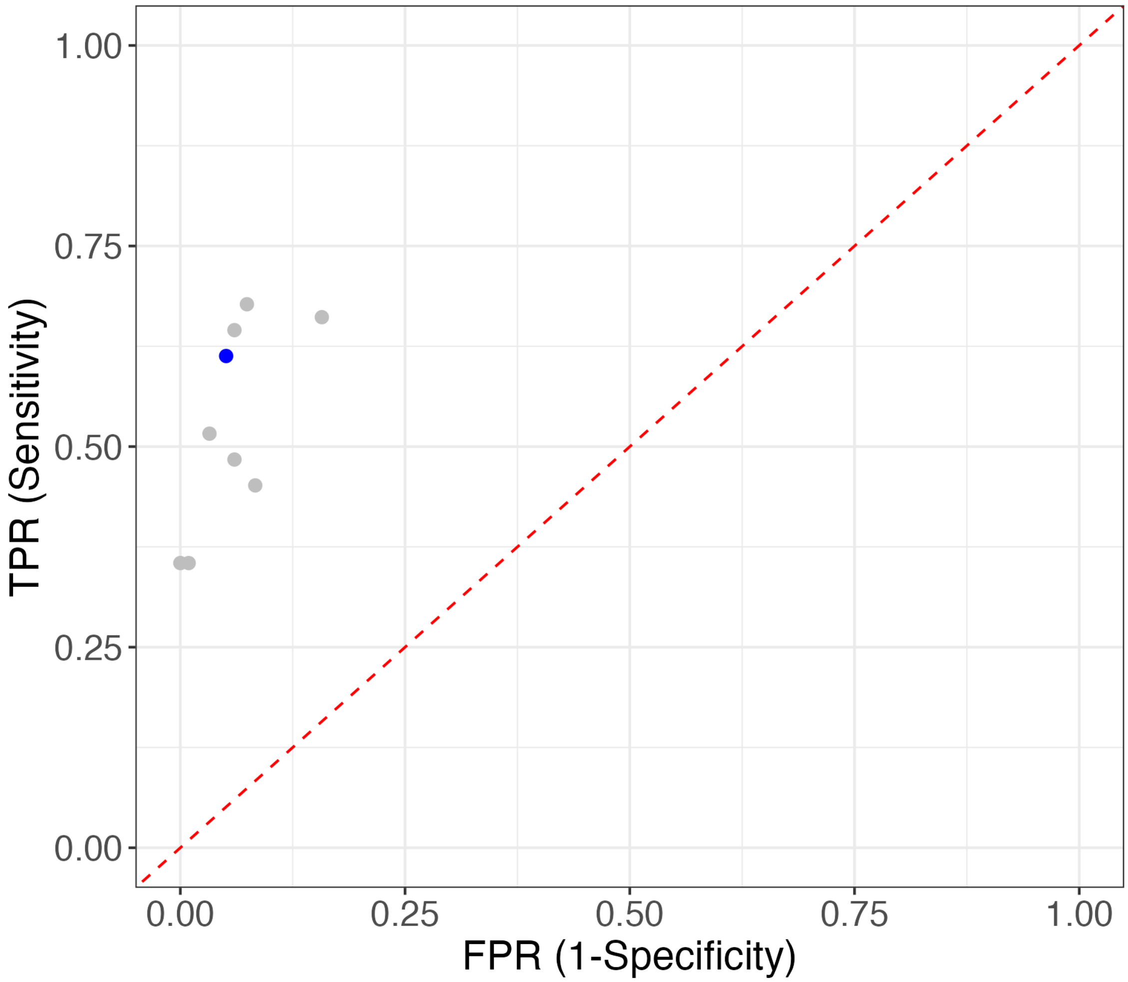
A.5.2 Discussion of potential biases from measurement of paper ideology

Our linking procedure, like most machine-linking procedures, is imperfect; it produces both false negatives and false positives. In this section, we discuss the possible biases that may arise due to our linking procedure.

Nonrepresentative sample: All scientists in our linked sample have donated to a political campaign. Given that the decision to donate to political campaigns is unlikely to be random, our sample may not be representative of the population of scientists. For example, these scientists may have stronger ideological beliefs than the median scientist. Additionally, scientists who donate may differ from non-donor scientists on non-political dimensions.

Noisy Measure: CF Score is an imperfect proxy for scientist ideology for two reasons. First, noise in the linking procedure creates some false positives. Second, our measure of paper slant only considers the slant of authors who have CF scores. The linked authors' influence on paper ideology might be overwhelmed by the influence of the rest of the team.

Figure S1: ROC curve comparing linking procedures



Notes: The Figure presents the receiver operating characteristic (ROC) of various linking procedures. Each observation is a linking procedure. The $x-$ axis presents the False Positive Rate and the $y-$ axis presents the True Positive Rate. The red-line represents the 45 degree line. The blue dot denotes our preferred linking procedure

We expect noise in our measure of scientist ideology to attenuate our estimates of alignment preferences towards zero. To evaluate the importance of attenuation bias, we ask whether our estimates of alignment preferences are larger over a set of papers where the CF score measure is less noisy. In particular, we allow our alignment measure to interact with an indicator for whether all authors of a paper were successfully linked. We present estimates from these specifications in Table S1. The results suggest that attenuation bias is a limitation of our study.

Table S1: Evaluating measurement error: estimates when all authors are linked

	Dependent variable:		
	Coverage Probability		
	(1)	(2)	(3)
Alignment	0.00058***	0.00051***	0.00042***
	(0.00011)	(0.00011)	(0.00011)
Alignment * All Authors Matched		0.00550***	0.00349***
		(0.00096)	(0.00096)
Field x Paper year F.E.	Yes	Yes	Yes
Outlet F.E.	Yes	Yes	Yes
Paper F.E.	Yes	Yes	Yes
Field x Outlet F.E.	No	No	Yes
Mean of Dep Variable	0.023	0.023	0.023
Observations	5,029,439	5,029,439	5,029,439
R^2	0.12418	0.12418	0.13084
Adjusted R^2	0.11809	0.11809	0.12426
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: This table presents estimates from the analyses described in the final paragraph of section A.5.2. Column (1) replicates the specification from column (1) of Table S5. Column (2) adds an interaction between Alignment and an indicator variable for whether all of a paper's authors were successfully linked to a CF score. We do not include the level effect of All Authors Matched because it is absorbed by the paper FEs. Column (3) adds field-outlet fixed effects.

B Descriptive statistics

B.1 Media outlets

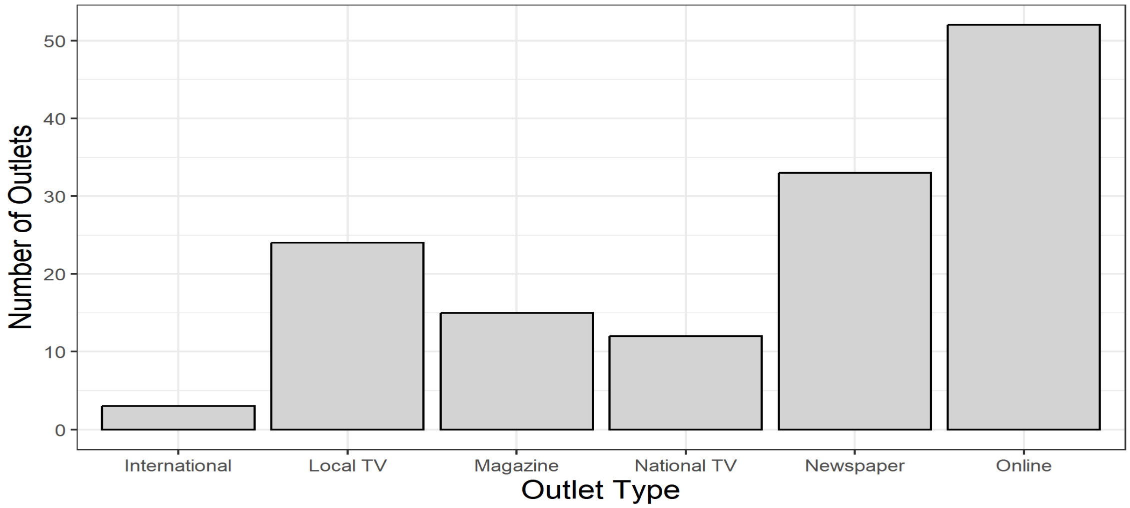
We begin by providing concrete examples of outlets in our sample. Table S2 lists the 10 outlets that covered the most articles. The largest outlets in our sample include large news aggregators like Yahoo! News, national newspapers like New York Times, national magazines like Forbes, and international outlets like BBC. In Figure S2, we present the number of outlets in our sample by outlet type. Most of the outlets are online news outlets and large national newspapers. In addition, our sample includes several local news outlets. The majority of the outlets in Table S2 are left-leaning. Figure S3 demonstrates that this is also true in the full sample.

Table S2: Examples of outlets in sample

Outlet	Scientific Articles Covered	BMA Slant
msn.com	23865	-0.08
news.yahoo.com	21101	0.05
nytimes.com	9949	-0.55
dailymail.co.uk	7295	0.29
forbes.com	5247	0.06
finance.yahoo.com	4944	0.08
washingtonpost.com	4898	-0.26
bbc.co.uk	3855	-0.33
edition.cnn.com	3815	-0.26
businessinsider.com	3693	-0.06

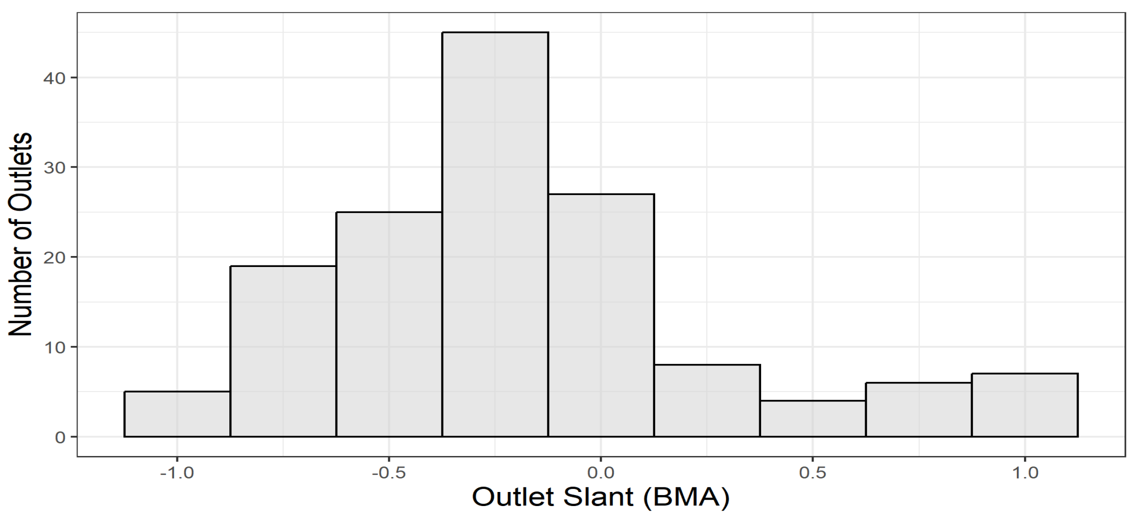
Notes: The table presents the top ten outlets by the number of scientific articles covered. The first column lists the outlet name. The second column lists the number of scientific articles covered. The third column lists the outlet slant as measured by the BMA score.

Figure S2: Distribution of outlet types



Notes: The figure presents the number of outlets by outlet type. The $x-$axis represents different outlet types and the $y-$axis represents the number of outlets of that type in our sample.

Figure S3: Distribution of outlet ideology



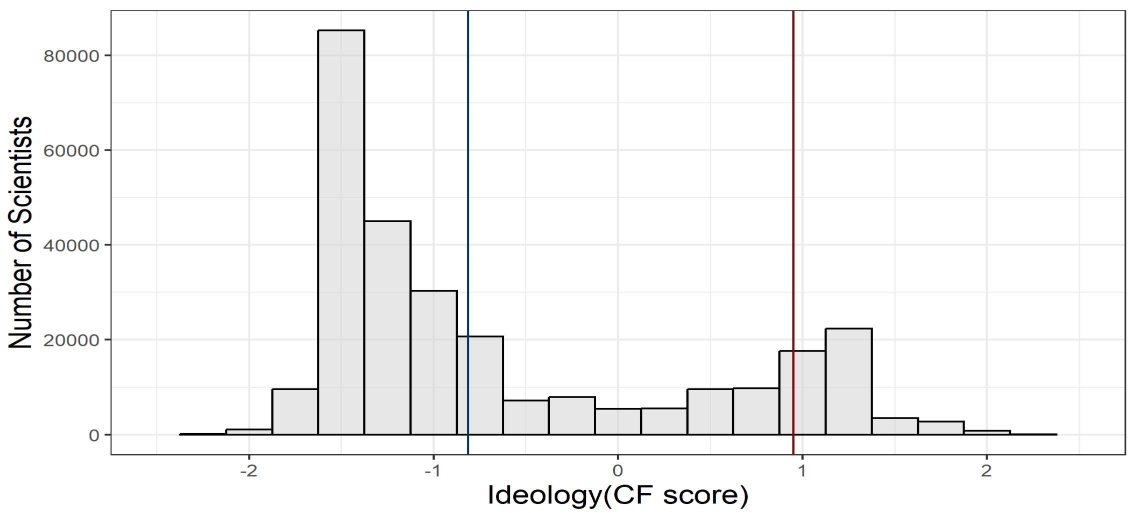
Notes: The figure presents the histogram of outlets by outlet slant. The $x-$axis represents the outlet slant as measured by the BMA score. The $y-$axis represents the number of outlets.

B.2 Ideology of scientists and fields of science

In Figure S4, we present the distribution of CF Scores of linked scientists. Most scientists in our sample are liberal. The ideological distribution of scientists in our sample is similar to the ideological distribution of all scientists in the US (4).

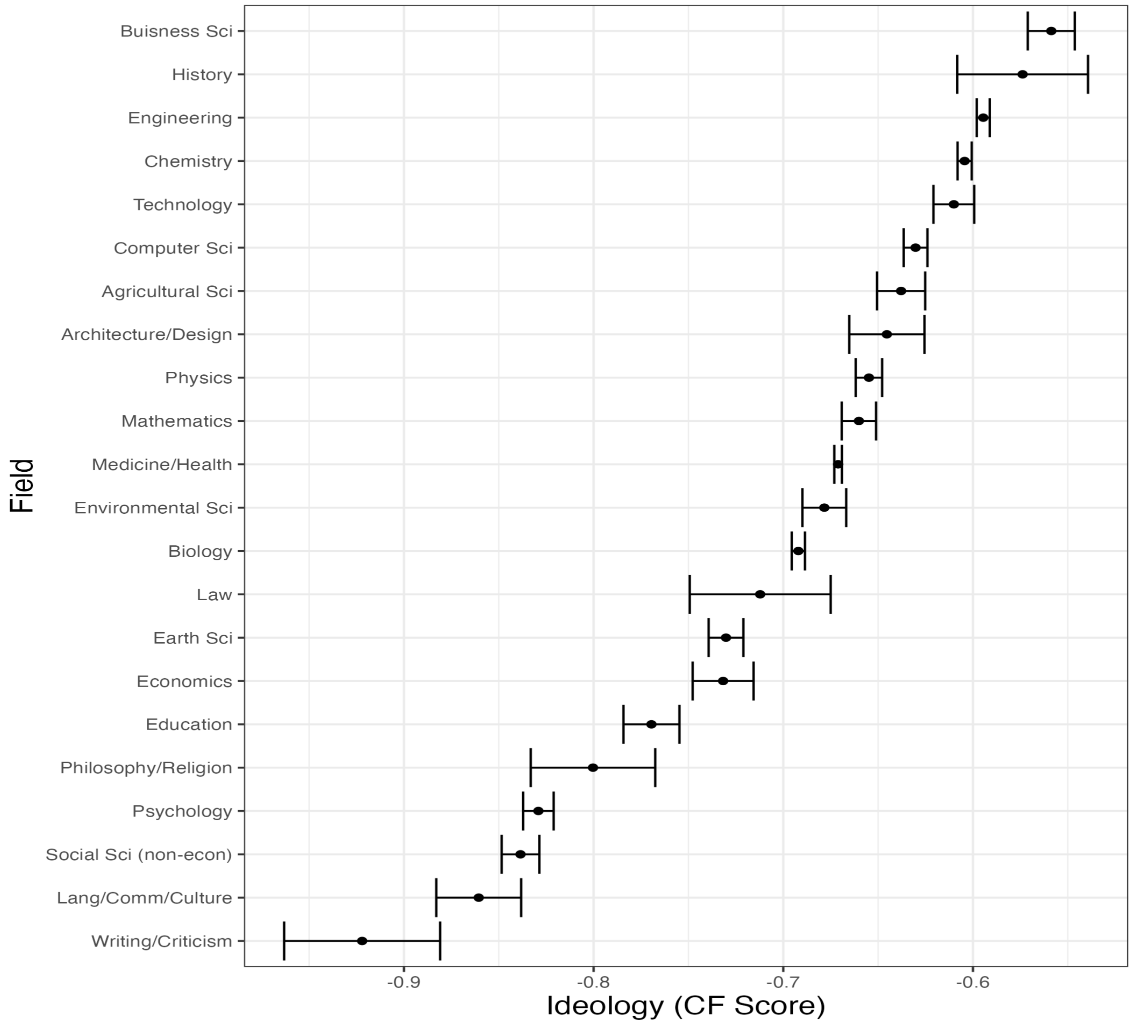
Figure S5 presents the mean ideology of papers across different scientific fields. Note that all fields, on average, are liberal-leaning. However, there is still considerable heterogeneity across fields. Fields like Business Science and Agriculture Science are more conservative relative to fields like Communications and Culture.

Figure S4: Distribution of scientist ideology



Notes: The figure presents the distribution of ideology for 259,826 linked scientists using our donations-based measure. The $x-$axis represents the scientist's ideology as measured by the CF Score. The $y-$axis represents the number of scientists. The blue and red vertical lines present the CF Score of the median Democratic and the median Republic senator, respectively.

Figure S5: Ideology of fields of science



Notes: The figure presents differences in mean ideology across fields of science. The $x-$axis represents ideology as measured by the CF Score. The $y-$axis represents different scientific fields. The points denoted the mean CF score and the lines denote the 95% confidence intervals for each field.

C Measuring ideological differences in science coverage

In this section, we define the GST measure from the main paper and present additional results.

C.1 Formal definitions

We loosely follow notation from (Gentzkow *et al.,* (5); henceforth GST). There are N_o outlets and N_p scientific papers. For each outlet-paper pair, we observe c_{op}, the number of times that outlet o cites paper p. For each outlet, we model c_{op} is the realization of a multinomial draw

$$c_{op} \sim MN(m_o, q_o) \tag{1}$$

where $m_o = \sum_p c_{op}$ denotes the total number of papers cited by outlet o and q_o is the vector of choice probabilities. We allow q_o to vary by outlet ideology, so $q_o = q^R$ for conservative outlets and $q_o = q^L$ for liberal outlets. An outlet is liberal if its BMA score is less than -0.2 and conservative if its BMA score is greater than 0.2. When computing the GST measure, we drop outlets with BMA scores between -0.2 and 0.2.

Our goal is to measure the divergence between q^R and q^L. Following GST, we motivate our divergence metric with the following thought experiment. Consider an observer who believes that a news outlet has an equal probability of being liberal or conservative. This observer sees a single scientific paper that was cited by the outlet, and knows the choice probability vectors q. If q^R and q^L are the same, then the observer has learned nothing about the outlet's type. By contrast, if conservative and liberal outlets cite completely different sets of articles, the observer has learned the outlet's type with certainty. In reality, the data will lie somewhere in between these extreme cases, and the divergence measure captures the extent to which the observer is able to update about the outlet type. Formally, we define an estimand π to be the posterior probability that this observer expects to assign to the outlet's true type.

$$\pi = \frac{1}{2} q^R \cdot \rho + \frac{1}{2} q^L \cdot (1 - \rho) \tag{2}$$

where the vector ρ captures the posteriors that the observer would assign to the outlet being conservative after observing each paper.

$$\rho_p = \frac{q_p^R}{q_p^R + q_p^D} \tag{3}$$

We estimate π from the data c_{op} on outlets citations of scientific papers using the leave-out estimator from GST. Let $Cons$ define the set of conservative outlets and Lib denote the set of liberal outlets. The estimator is

$$\hat{\pi} = \frac{1}{2}\frac{1}{|Cons|}\sum_{o\in Cons}\hat{q}_o \cdot \hat{\rho}_{-p} + \frac{1}{2}\frac{1}{|Cons|}\sum_{o\in Cons}\hat{q}_o \cdot \hat{\rho}_{-p} \tag{4}$$

where $\hat{q}^G$ is the empirical paper frequency for outlet o

$$\hat{q}_p^C = \frac{\sum_{o\in Cons} c_{op}}{\sum_{o\in Cons}\sum_{p'} c_{op'}} \tag{5}$$

$$\hat{q}_p^L = \frac{\sum_{o\in Lib} c_{op}}{\sum_{o\in Lib}\sum_{p'} c_{op'}} \tag{6}$$

and $\hat{\rho}_{-p}$ is the leave-out empirical analog of ρ, i.e, the fraction of paper p's cites that come from conservative outlets among all outlets except o.[4]

Whenever we discuss GST measures, we are referring to estimates using the estimator in equation (4).

C.2 Estimates

Table S3 presents GST measures for several subsamples of the data. We discuss the interpretation of these estimates in the main paper.

[4] We discard from the sample all papers that are cited by only one news outlet.

Table S3: GST measures of differences in citation patterns

Sample	Group Comparison	Estimate	N Outlets	N Citations
Main	Lib vs. Cons	0.567	103	142847
Main	Online vs. Print Newspapers	0.536	86	181382
Main	TV vs. Print Newspapers	0.561	74	90967
No CF Score	Lib vs. Cons	0.575	103	73832
Yes CF Score	Lib vs. Cons	0.560	103	69015
Pre-2020	Lib vs. Cons	0.574	103	114387
2020, Not COVID	Lib vs. Cons	0.560	90	17146
2020, Yes COVID	Lib vs. Cons	0.594	88	9884
Medicine/Health	Lib vs. Cons	0.559	103	75402
Biology	Lib vs. Cons	0.584	101	14550
Psychology	Lib vs. Cons	0.595	97	9310
Earth Sci	Lib vs. Cons	0.599	98	9151
Social Sci (non-econ)	Lib vs. Cons	0.576	95	4897
Physics	Lib vs. Cons	0.568	81	4593
Engineering	Lib vs. Cons	0.602	99	4093
Environmental Sci	Lib vs. Cons	0.551	92	3566
Chemistry	Lib vs. Cons	0.575	91	2537
Economics	Lib vs. Cons	0.588	92	1837
Computer Sci	Lib vs. Cons	0.593	85	1652
History	Lib vs. Cons	0.607	77	1593
Agricultural Sci	Lib vs. Cons	0.540	81	961

Notes: The table presents GST measures for various subsamples. We truncate citation counts c_{op} so that they are either zero (if o does not cite p) or one (if o cites p). The "Group Comparison" column explains the definition of outlet type q_o for that row. So in the first row, we report a measure of the divergence in citations between liberal and conservative outlets, while in second row, we report a a measure of the divergence in citations between online outlets and print newspapers. Liberal vs. conservative outlet types are defined in the text. Online, Print Newspaper, and TV outlet types are manually coded. Field categorizations are from Dimensions, using ANZRC codes. COVID papers are papers that appear in the CORD (6) database.

C.3 Randomization tests

We conduct randomization tests to establish statistical significance. The idea is to compare the observed GST measure to simulated GST measures under the null hypothesis that $q^R = q^D$. In our simulations, we randomly assign parties to outlets, so the only differences between $\hat{q}^R$ and $\hat{q}^D$ in the simulations are due to sampling error. We conduct 100 such simulations for each estimate in table S3. We visualize our procedure for the main estimate in Figure S6, and present simulation results for all estimates in Table S4.

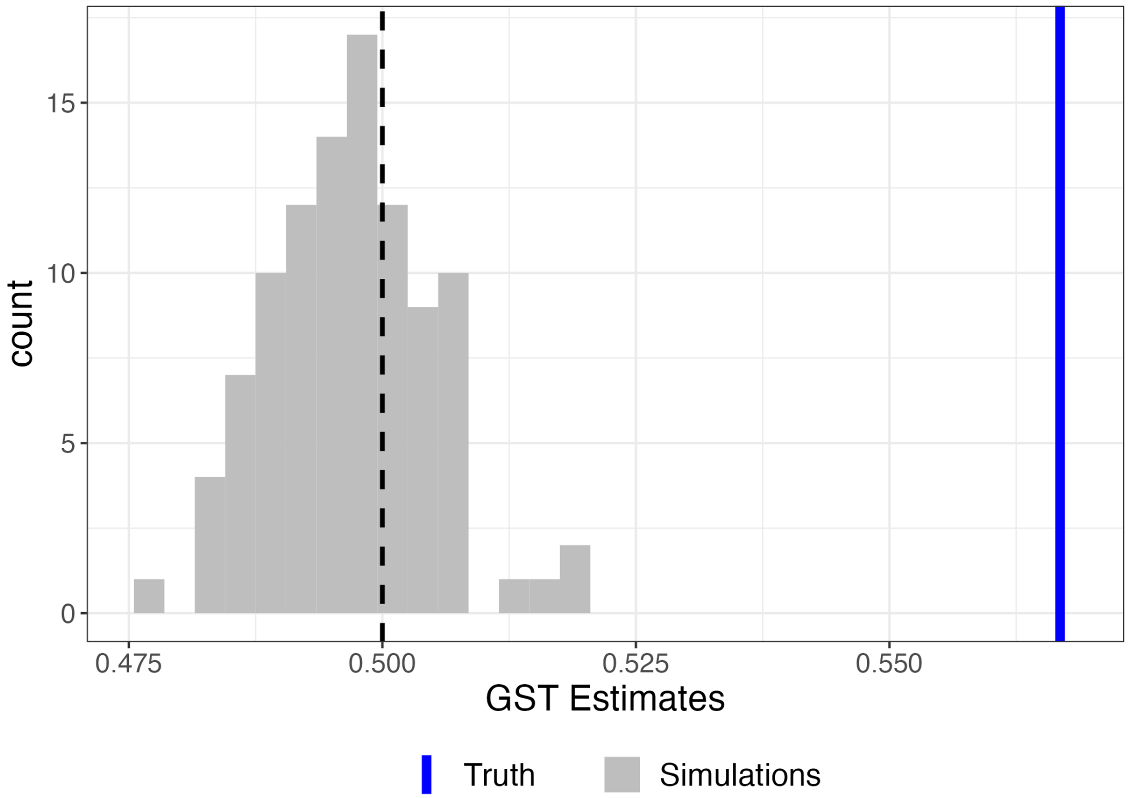


Figure S6: Visualizing GST randomization tests

Notes: We conduct 100 simulations of our main GST specification (the top row of table S3) in which we randomly assign party to outlet. The simulated estimates are represented in gray, while the true value is represented with a blue line. The true estimate is greater than any of the simulated estimates, which indicates statistical significance.

Table S4: GST Randomization inference results

Sample	Group Comparison	Estimate	Sim 95th pctile	Sim p-value
Main	Lib vs. Cons	0.567	0.510	0.00
Main	Online vs. Print Newspapers	0.536	0.510	0.00
Main	TV vs. Print Newspapers	0.561	0.514	0.00
No CF Score	Lib vs. Cons	0.575	0.517	0.00
Yes CF Score	Lib vs. Cons	0.560	0.512	0.00
Pre-2020	Lib vs. Cons	0.574	0.517	0.00
2020, Not COVID	Lib vs. Cons	0.560	0.513	0.00
2020, Yes COVID	Lib vs. Cons	0.594	0.533	0.00
Medicine/Health	Lib vs. Cons	0.559	0.508	0.00
Biology	Lib vs. Cons	0.584	0.515	0.00
Psychology	Lib vs. Cons	0.595	0.524	0.00
Earth Sci	Lib vs. Cons	0.599	0.527	0.00
Social Sci (non-econ)	Lib vs. Cons	0.576	0.532	0.00
Physics	Lib vs. Cons	0.568	0.538	0.00
Engineering	Lib vs. Cons	0.602	0.538	0.00
Environmental Sci	Lib vs. Cons	0.551	0.528	0.01
Chemistry	Lib vs. Cons	0.575	0.533	0.01
Economics	Lib vs. Cons	0.588	0.536	0.01
Computer Sci	Lib vs. Cons	0.593	0.541	0.00
History	Lib vs. Cons	0.607	0.535	0.00
Agricultural Sci	Lib vs. Cons	0.540	0.543	0.06

Notes: For each GST specification, we conduct 100 simulations in which we randomly assign party to outlet. The simulated p-value is the fraction of simulations for which the true estimate is less than the simulated estimate.

D Linear probability models

We estimate models of the following type on the set of all **o**utlet-**p**aper pairs (o, p):

$$Y_{op} = \beta a(I_o, I_p) + \gamma X_{op} + \epsilon_{op}\,, \tag{7}$$

where Y_{op} is an indicator for whether the paper was covered by the outlet at least once between 2015-2020, $a(I_o, I_p)$ is the ideological alignment between outlet ideology I_o and paper ideology I_p, X_{op} represents controls, and ϵ_{op} represents idiosyncratic errors. Our primary measure of alignment $a(I_o, I_p)$ is the absolute value of the difference in ideology scores for the paper and outlet $a(I_o, I_p) = |I_o - I_p|$. To aid with interpretation, we standardize the measure so that a coefficient corresponds to the effect of a 1 SD increase in alignment. Our results are robust to alternative alignment measures (see Section E.1). The parameter β captures the association between alignment and coverage probability, conditional on controls X_{op}. In all specifications, our controls X_{op} include fixed effects that flexibly control for outlet-specific or paper-specific features.

We present estimates in Table S5. Column (1) is our baseline specification. Columns (2) and (3) control for two possible confounders by which an association may arise between alignment and citation probability may arise. In column (2), we control for outlets' propensities to cite papers within different fields of science, and in column (3), we control for outlets' different preferences over observable markers of academic quality. Column (4) controls for both of these confounders.

Table S5: Parametric estimates of alignment association (LPMs)

	Dependent variable:			
	Coverage Probability			
	(1)	(2)	(3)	(4)
Alignment	0.00058***	0.00046***	0.00054***	0.00042***
	(0.00011)	(0.00011)	(0.00011)	(0.00011)
Field x Paper year F.E.	Yes	Yes	Yes	Yes
Paper F.E.	Yes	Yes	Yes	Yes
Outlet F.E.	Yes	No	Yes	No
Field x Outlet F.E.	No	Yes	No	Yes
Outlet F.E. x Paper Quality	No	No	Yes	Yes
Mean of Dep Variable	0.023	0.0231	0.023	0.0231
Observations	5,029,439	5,029,439	4,964,177	4,964,177
R^2	0.12418	0.13084	0.12924	0.13585
Adjusted R^2	0.11809	0.12425	0.12316	0.12927
Note:				*p<0.1; **p<0.05; ***p<0.01

Notes: The table presents the parameter estimates of the linear probability model defined in Equation 7 for different sets of controls. An observation is an outlet-paper pair. The parameter of interest is β, which captures the extent to which ideological alignment a_{op} between an outlet and a paper is associated with coverage probability. Alignment $a(I_o, I_p)$ is the standardized absolute value of the difference in the ideology of the paper and outlet. We only include papers that were cited by at least one outlet. In Column (1) we included field-year, paper, and outlet fixed effects. In column (2), we add field-outlet fixed effects, and in column (3) we interact outlet fixed effects with the paper quality measure. In column (4), we include both field-outlet fixed effects and the outlet fixed effect interacted with the paper quality index.

E Robustness checks

E.1 Alternate alignment measures

In our main analysis, we define alignment between an outlet and a paper as the absolute value of the difference in the ideology $a(I_o, I_p) = |I_o - I_p|$. A potential measurement issue is that the BMA score and CF Scores are on different scales. Even though 0 represents the ideological center for both measures (and so the signs of the measures are consistent), the magnitudes are incommensurate. To assuage these concerns, we employ two other measures of alignment.

Our first alternate measure compares relative ideological positions. For each outlet, we estimate the quantile of their ideology in the distribution outlet ideologies $q(I_o)$. We apply the same procedure for scientific papers $q(I_p)$.[5] Our relative alignment measure is the absolute value of the difference between these quantiles $a_q(I_o, I_p) = |q(I_o) - q(I_p)|$. By construction, the scales of the quantile measures are commensurate. In this measure, an outlet is perfectly aligned with a paper if both fall at the same point in their corresponding ideological distributions. Note that the ideological distribution of scientific ideology is left-skewed compared to that of outlet ideology. Therefore, a paper that is as conservative as an outlet on the relative scale will be more liberal on the absolute scale. We emphasize that there is no simple mapping between these two alignment measures as they capture different notions of alignment. Our second measure is a binary measure of alignment $a_b(I_o, I_p) = 1\{\text{sign}(I_o) = \text{sign}(I_a)\}$: an outlet and paper are aligned if they are both liberal (conservative); otherwise, they are not aligned. This measure only relies on the sign of the ideology and not the magnitude. This measure is invariant to any scale changes in the ideology measure. However, there is a loss of information with this measure, since it is a coarser measure of alignment.

Table S6 presents estimates of alignment preferences using these alternative measures. The regression specification is that defined in Equation 7. We include field-year and paper fixed effects in all specifications. We also include either outlet or outlet-field fixed effects. Across all regression, we consistently find positive and quantitatively meaningful alignment preferences. These results are significant at all conventional levels when we use the relative alignment measure a_q. The alignment preferences are both quantitatively smaller and more imprecisely estimated with our binary alignment measure a_b. The smaller magnitudes suggest that fine ideological differences between an outlet and paper influence coverage propensity, not just whether they are on the same side of the political aisle. The estimates for the binary measure are less precisely estimated in part because once we account for paper, field, and outlet fixed effects there is very little variation in the binary measure of alignment. This is because, in many fields, most papers are liberal.

In summary, these exercises highlight that the presence of alignment preferences is not driven by any particular definition of alignment. As one would expect, the interpretation and magnitude of these alignment preferences do vary across these different definitions.

[5] Formally, $\mathbb{P}(I > I_o) = q(I_o)$, where I is distributed according to the ideology distribution of outlets.

Table S6: Alignment associations using alternate alignment Measures

	Dependent variable:			
	Coverage Probability			
	(1)	(2)	(3)	(4)
Alignment (Relative Scale)	0.00277***	0.00203***		
	(0.00030)	(0.00030)		
Alignment (Binary Scale)			0.00019**	0.00014
			(0.00009)	(0.00009)
Field x Paper year F.E.	Yes	Yes	Yes	Yes
Outlet F.E.	Yes	Yes	Yes	Yes
Paper F.E.	Yes	Yes	Yes	Yes
Field x Outlet F.E.	No	Yes	No	Yes
Mean of Dep Variable	0.023	0.023	0.023	0.023
Observations	5,029,439	5,029,439	5,029,439	5,029,439
R^2	0.12419	0.13084	0.12417	0.13083
Adjusted R^2	0.11810	0.12421	0.11808	0.12420
Note:		*p<0.1; **p<0.05; ***p<0.01		

Notes: The table presents the parameter estimates of the linear probability model defined in Equation 7 for different sets of controls and measures of alignment. Each observation in the model is an outlet-paper pair. The parameter of interest is β which captures the extent to which ideological alignment a_{op} between an outlet and a paper is associated with coverage probability. In columns (1) and (2), alignment $a_q(I_o, I_p)$ is the absolute value of the difference in the ideology quantile of the paper and outlet and in columns (3) and (4), alignment $a_b(I_o, I_p)$ is an indicator which is equal to 1 if both the paper and outlet are Liberal (conservative). We only include papers that were cited by at least one outlet. All columns include field-year and paper fixed effects. Columns (1) and (3) include outlet fixed effects. Columns (2) and (4) include outlet-field fixed effects.

E.2 Alternate Linking Procedures

In this section, we show that our main result is robust to our choice of linking procedure. In Table S7, we estimate alignment preferences for each of five different linking procedures, including the procedure we used in the main paper. The results from our preferred procedure are given in column (1). Each of the other four columns uses a different procedure.

The procedure for column (2) includes all scientists who had an exact name match in the expanded donor sample (if a scientist has multiple exact name matches in the expanded donor sample, we assign the scientist the mean CF score of their matches). The procedure for column (3) includes all scientists who had an exact name match in the strict donor sample. The procedure for column (4) includes all scientists who had a unique match with a name-distance less than or equal to one in the strict donor sample. The procedure for column (5) includes all scientists who had any match in the strict donor sample, and also all scientists who had a unique match in the expanded donor sample.

In all columns, our alignment preference estimate is positive and statistically significant, which demonstrates that our main result was not driven by the idiosyncrasies of our chosen linking procedure.

Table S7: Alignment preferences with alternate linking procedures

	Dependent variable:				
	Coverage Probability				
	(1)	(2)	(3)	(4)	(5)
Alignment	0.00058***	0.00031***	0.00059***	0.00036**	0.00042***
	(0.00011)	(0.00011)	(0.00020)	(0.00017)	(0.00011)
Field x Paper year F.E.	Yes	Yes	Yes	Yes	Yes
Outlet F.E.	Yes	Yes	Yes	Yes	Yes
Paper F.E.	Yes	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.023	0.023	0.024	0.023	0.023
Observations	5,029,439	4,652,619	2,869,638	3,549,590	5,961,968
R^2	0.12418	0.12213	0.12151	0.12439	0.12348
Adjusted R^2	0.11809	0.11602	0.11536	0.11828	0.11739
Note:					*p<0.1; **p<0.05; ***p<0.01

Notes: This table tests the robustness of our main alignment association estimates to alternative linking procedures. We present parameter estimates of the model defined in equation 7. Column (1) replicates the result from column (1) of Table S5 in the main paper. Columns (2) through (5) each use different procedures, which are described in section E.2.

E.3 Parametric quality analysis

We examine how coverage probability depends on alignment and different quality measures. We estimate the following linear LPMs

$$Y_{op} = \beta a(I_o, I_p) + \alpha q_p + \gamma X_{op} + \epsilon_{op} \,, \tag{8}$$

where Y_{op} is an indicator for whether the paper was covered by the outlet at least once between 2015-2020, $a(I_o, I_p)$ is the ideological alignment between outlet ideology I_o and paper ideology I_p, q_p is the quality of paper p, X_{op} represents controls, and ϵ_{op} represents idiosyncratic errors. We include field-year fixed effects and field-outlet fixed effects as controls. We examine four measures of quality: number of academic citations, amount of research grants, number of patent citations, and journal impact.

The results are presented in Table S8. Across all specifications, we find a positive association between alignment and coverage probability that is comparable to our main estimates. For each quality measure, we find that coverage probability increases in quality. The association between quality measures and coverage probability is stronger than that of alignment and quality measures. This result reaffirms that quality affects coverage probability much more than alignment.

Table S8: Parametric Estimates of Alignment and Quality Association

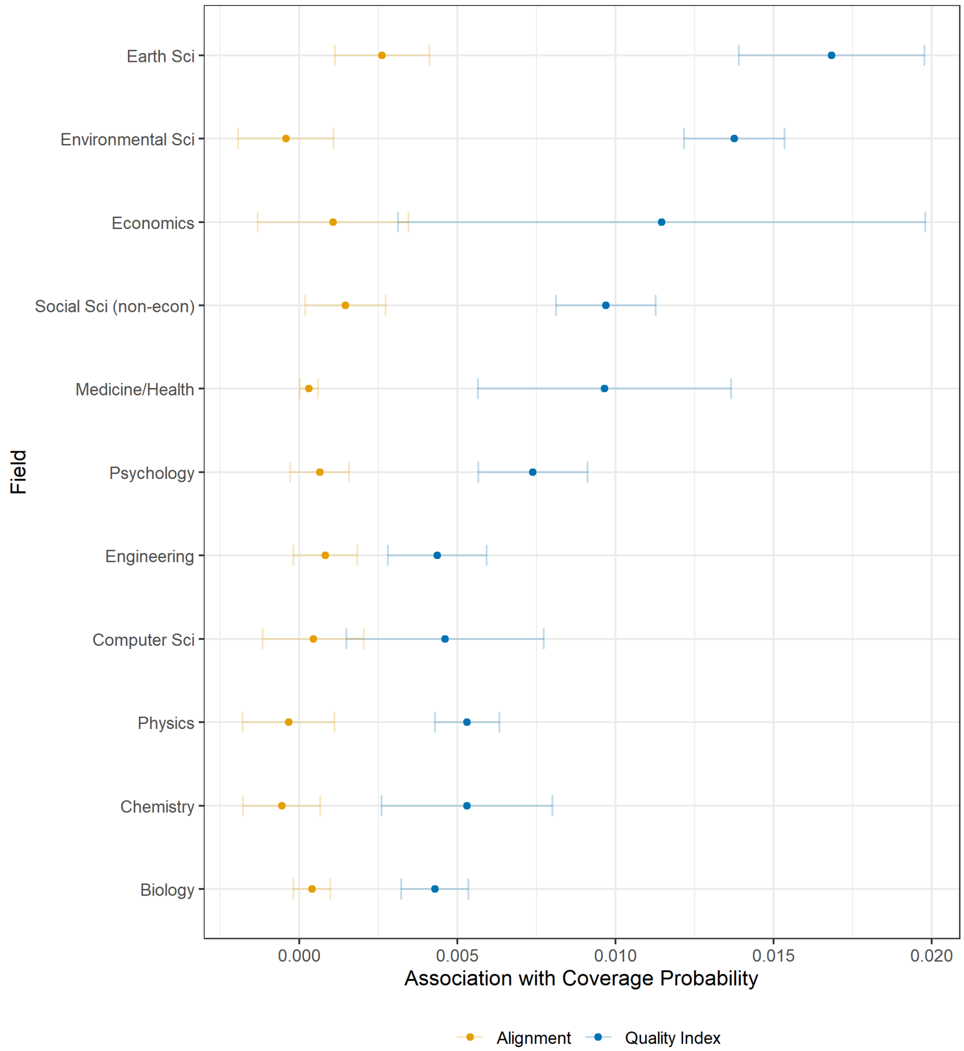
	Dependent variable:			
	Coverage Probability			
	(1)	(2)	(3)	(4)
Alignment	0.00009	0.00036***	0.00036***	0.00015**
	(0.00007)	(0.00007)	(0.00007)	(0.00007)
Academic Citations (1000s)	0.02061***			
	(0.00016)			
Research Grants (log)		0.00007***		
		(0.00001)		
Patents (log)			0.00286***	
			(0.00027)	
Journal Impact				0.00002***
				(0.0000002)
Field x Year F.E.	Yes	Yes	Yes	Yes
Field x Outlet F.E.	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.023	0.023	0.023	0.023
Observations	5,029,439	5,029,439	5,029,439	4,969,725
R^2	0.07909	0.08269	0.08270	0.08411
Adjusted R^2	0.07903	0.08203	0.08203	0.08343
Note:				*p<0.1; **p<0.05; ***p<0.01

Notes: The table presents the parameter estimates of the linear probability model defined in Equation 8 for different quality measures. Each observation in the model is an outlet-paper pair. Alignment $a_q(I_o, I_p)$ is the standardized absolute value of the difference in the ideology of the paper and outlet. The quality measures are number of academic citations, amount of research grants, number of patent citations, and journal impact ($h-$ index) for columns (1), (2), (3), and (4), respectively. We only include papers that were cited by at least one outlet. All columns include field-year and filed-outlet fixed effects.

E.4 Quality association by field of science

In Figure S7 we present the association between coverage propensity and alignment alongside the association between coverage propensity and the quality index across fields of science. For each field, these associations are estimated simultaneously using an LPM as in Equation 8. The figure shows that quality is a stronger predictor of media coverage than ideological alignment, even for fields such as Earth Science where alignment association is the strongest.

Figure S7: Heterogeneity in alignment and quality associations by field of science



Notes: The figure presents the relationship between alignment and quality with media coverage for different fields of science. The $x-$axis represents a field of science. The $y-$axis presents the association between coverage propensity and alignment as well as quality, denoted by different colors. The association is estimated using (separate) linear probability models. The alignment measure and quality index are normalized to have a mean of 0 and a standard deviation of 1.

References

[1] E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on facebook, *Science* **348**, 1130–1132 (2015).

[2] C. Herzog, D. Hook, S. Konkiel, Dimensions: Bringing down barriers between scientometricians and data, *Quantitative Science Studies* **1**, 387–395 (2020).

[3] A. Bonica, Database on ideology, money in politics, and elections, *Public version 2.0 [Computer file]. Stanford, CA: Stanford University Libraries.* (2016).

[4] A. Bonica, Mapping the ideological marketplace, *Am. J. Pol. Sci.* **58**, 367–386 (2014).

[5] M. Gentzkow, J. M. Shapiro, M. Taddy, Measuring group differences in high-dimensional choices: method and application to congressional speech, *Econometrica* **87**, 1307–1340 (2019).

[6] L. L. Wang, *et al.*, Cord-19: The covid-19 open research dataset, *arXiv:2004.10706* (2020).