

Millions of online book co-purchases reveal partisan differences in the consumption of science

Feng Shi^{1†}, Yongren Shi^{2†}, Fedor A. Dokshin³, James A. Evans^{1,4*} and Michael W. Macy^{3*}

Passionate disagreements about climate change, stem cell research and evolution raise concerns that science has become a new battlefield in the culture wars. We used data derived from millions of online co-purchases as a behavioural indicator for whether shared interest in science bridges political differences or selective attention reinforces existing divisions. Findings reveal partisan preferences both within and across scientific disciplines. Across fields, customers for liberal or 'blue' political books prefer basic science (for example, physics, astronomy and zoology), whereas conservative or 'red' customers prefer applied and commercial science (for example, criminology, medicine and geophysics). Within disciplines, 'red' books tend to be co-purchased with a narrower subset of science books on the periphery of the discipline. We conclude that the political left and right share an interest in science in general, but not science in particular. This underscores the need for research into remedies that can attenuate selective exposure to 'convenient truth', renew the capacity for science to inform political debate and temper partisan passions.

In its quest for an objective understanding of the world¹, modern science has practised two distinct forms of political neutrality: as an apolitical 'separate sphere' detached from ideological debates, and as a 'public sphere' relevant to political issues but with balanced political engagement that aids reasoned deliberation and deference to evidence^{2–5}. Recent surveys support the view that science contributes not only to human knowledge but also to social integration, both as a voice of reason and also as a shared value. Joint surveys conducted by the American Association for the Advancement of Science (AAAS) and the Pew Research Center in 2009 and 2014 found that science remains near the top in public rankings of professions, well above that of clergy, despite the prevalence of liberals among scientists^{6–8}. Although nearly two-thirds of respondents question evolution, even those who see conflict with issues of personal faith overwhelmingly support scientific contributions to public well-being (67%). In a highly polarized electorate, these responses invite reassurance that science continues to command political deference as a voice of reason and bridges the partisan divide. We may disagree on emotionally charged social issues, but at least we can agree on science.

Political and cultural polarization within the United States, however, raises questions about the validity of this interpretation⁹. A less comforting possibility is that verbal survey responses may simply echo an Enlightenment commitment to value-free scientific inquiry that masks underlying scepticism about science. In recent years, conservative politicians and pundits have challenged scientific positions on evolution, cosmology, climate change and the

perceived liberal bias in policies advocated by social scientists. For example, the conservative-funded scientific counter-movement in climate change research suggests the possibility of politically driven scientific polarization^{10,11}. When science becomes politicized, partisans tend to cast doubt on scientific consensus through questioning its inherent uncertainty^{12–14}. This process is manifest not only in conservative resistance to climate change, but in historically liberal resistance to consensus over the positive benefits of genetically modified organisms, vaccination, nuclear power and the safe storage of nuclear waste¹⁵.

Survey data show little overall change in public confidence in science since 1970, but beneath the surface there is a marked shift: conservatives in the Vietnam era were more confident in science than liberals, but today that pattern has reversed¹⁶ (Supplementary Fig. 1). Does public exposure to science play an integrative role by encouraging and informing empirical validation? Or has selective attention instead reinforced the 'Big Sort' of American politics^{17–19} — the tendency to cluster in like-minded communities?

Much previous research has used surveys to investigate political alignments of the producers of science (with a few exceptions^{20,21}). We focus instead on the consumers of science, using online co-purchases of books on science and politics as a behavioural indication of preferences held by customers who 'vote with their pocket-book', in contrast to survey responses that are costless. Surveys measure what researchers think is important, not what respondents care about, whereas online consumers can register their preferences by purchasing books on any topic they choose. Retrospective self-reports are vulnerable to lapses of memory, whereas online sellers track every purchase. Survey responses are difficult to align across instruments that ask different questions and ask questions differently, whereas books from different stores can be classified using consistent typologies (for example the Library of Congress). Surveys are vulnerable to response bias from participants reluctant to reveal views regarded as politically incorrect; books purchased online arrive cloaked in cardboard. Finally, although surveys can use stratified random samples to generalize results to the underlying population, which is not possible with data from a convenience sample, rates of non-response are rising in landline-administered surveys, which raises concerns about their external validity²².

We addressed concerns about generalizability in two ways — by replicating our analysis using two independent samples of purchasing behaviour from two online merchants (Amazon and Barnes & Noble), and also by the size of these samples, collectively comprising hundreds of millions of online customers, including members of hidden populations (such as those without landlines) who may be undercounted in surveys based on at most a few thousand respondents.

¹Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, USA. ²Yale Institute for Network Science, Yale University, 17 Hillhouse Avenue, New Haven, Connecticut 06511, USA. ³Department of Sociology, Cornell University, 323 Uris Hall, Ithaca, New York 14853, USA.

⁴Department of Sociology, University of Chicago, 1126 East 59th Street, Chicago, Illinois 60637, USA. [†]These authors contributed equally to this work.

*e-mail: jevans@uchicago.edu; mwm14@cornell.edu

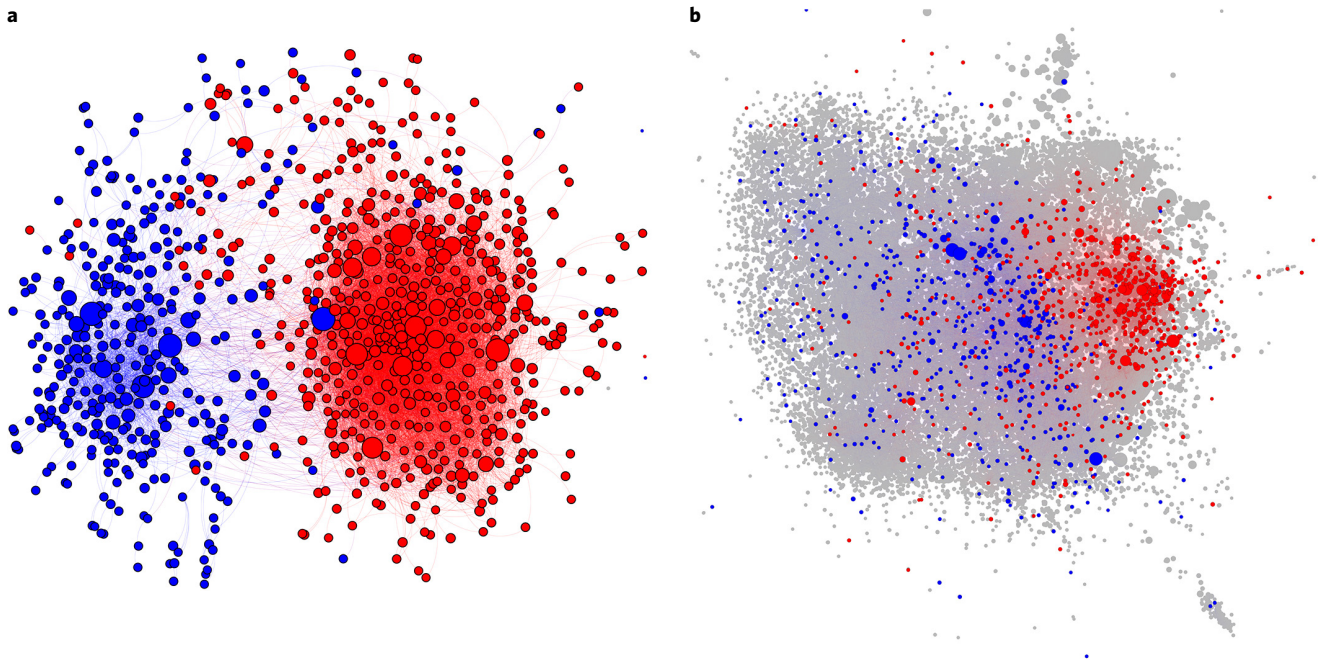


Figure 1 | Visualization of the co-purchase network among liberal, conservative and scientific books. **a**, Links between 583 liberal (blue) and 673 conservative (red) books. **b**, Links between these books and science (grey) books. As shown in **a**, 97.2% of red books linked to other reds and 93.7% of blue books linked to other blues. A small number of books were more likely to be co-purchased with books of a different colour. We subjected these blinded books to additional judges and found that the original codings were nearly all correct. A number of red ‘orphans’ were written by moderate Republicans critical of the religious right while blue ‘orphans’ were written by progressive community organizers such as Saul Alinsky (*Rules for Radicals*⁴⁰), later rediscovered by the Tea Party who reference their effective ethnic blue-collar organizing tactics. In **b**, the broader distribution of blue-linked science books indicates that readers of blue books have broader disciplinary interest and co-purchased books more centrally located in the network of co-purchased science books.

Our approach also differs from survey methods in the unit of analysis. Individuals are the units in surveys, but online retailers do not provide access to individual customer behaviour. Instead, we use individual books as the unit of analysis in constructing a co-purchase network. Bipartite network analysis has been widely used in research on co-citation and co-author networks^{23–27} where individual information is aggregated to provide connections between scientific papers. What is lost in the absence of individual-level responses for a small sample is compensated for with massive numbers of observations we can use to detect population-level patterns generated by individual behaviour. Here we extend bipartite network analysis to the study of online co-purchase behaviour among a diverse population of tens of millions of consumers. These data provide an unprecedented opportunity to study the entire popular audience for science in ways that are not possible using traditional methods.

These data do not speak to the partisan alignment of scientists, the policy relevance of scientific research or the political polarization of science as an institution. Nor do we address the political preferences of science’s consumers. Rather, our attention is focused exclusively on the science preferences of those who purchase liberal and conservative political books, a group whose science preferences could differ from those who do not shop for books online or who shop for science books but not for politics. Within that constraint of available data, we ask to what extent purchasers of political books are also interested in science, and in what parts of science they are most interested. A shared interest in science might provide a bridge across partisan divisions, whereas selective attention to ‘convenient truths’ risks reinforcement of existing political identities.

To find out, we constructed two undirected co-purchase networks of books from the American domain of the world’s two largest online book stores, following an approach pioneered by Valdis Krebs^{28–30}. Up to 100 unranked books are listed on each book page under the heading “Customers Who Bought This Item

Also Bought”. According to previous work³¹, these recommendations are based on a disproportionate number of customers who bought both books and are designed exclusively to identify those books (up to 100) that a customer is most likely to be interested in also purchasing. We compared results derived from two different bookstores — Amazon, and Barnes & Noble — each with its own co-purchasing algorithm. No important differences were found (see Supplementary Information).

Differences in the relative popularity of two books A and B could cause B to be included in the top 100 list for A while A does not make it into the list for B. Co-purchase behaviour is inherently undirected at the level of an individual customer’s multiple purchases, however, as the purchase of book A cannot be said to cause an individual to purchase B any more than the converse. We therefore ignore direction and reciprocation and define an undirected co-purchase link between two books as the level of bi-directed co-purchasing required to trigger a co-purchase listing in either direction.

Beginning with several seed books, we collected data recursively by tracing co-purchase links, iterating the search until no new titles could be identified. In total, we collected 26,467,385 co-purchase links among 1,303,504 books from amazon.com after consolidating multiple editions. (For details, see Methods. These counts refer to the number of titles, not purchases; we use ‘books’ to refer to titles and never to the physical objects.) From this collection, we identified three groups: political books, science books, and non-science books. From 3,530 politically relevant titles, two independent coders (with a third as tie-breaker) identified 673 conservative (‘red’) books, 583 liberal (‘blue’) books, and discarded 2,274 indeterminate books. The ‘indeterminate’ books include some that are centrist, and these could be used to investigate different interests in science held by moderates and those with more extreme views (whether red or blue). That possibility falls outside the scope of the present study, which is limited to comparisons of red and blue co-purchases.

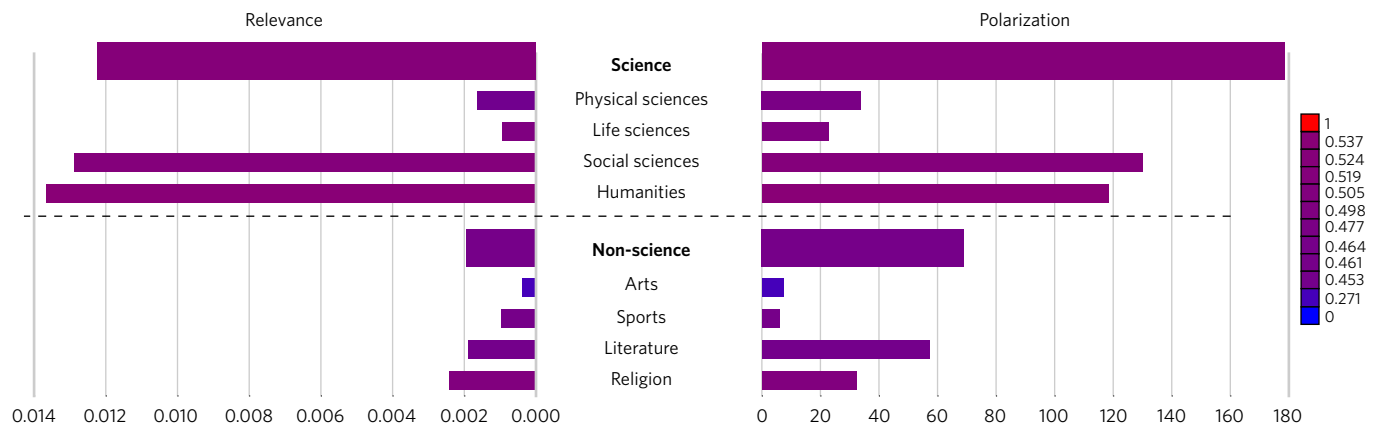


Figure 2 | Comparisons of political relevance, polarization and alignment between science and non-science books. Colour indicates alignment from liberal (blue) to conservative (red), corresponding to the relative interest in the topic by liberals versus conservatives. The widths of the bars (x axis) in the left panel indicate political relevance of topics of science and non-science books, and those in the right panel indicate their political polarization. Science books attract equal interest from liberals and conservatives but are more politically relevant and polarized than non-science books, largely owing to the social sciences and humanities, while the physical and life sciences are similar to non-science overall. Books on the performing arts are most aligned with liberals and, along with sports, have low political relevance and polarization compared with literature, religion and science.

As an additional validation, we imputed red and blue codings based on the relative number of links to other red and blue books and compared these with the hand codings. Over 96% were in agreement. The network among blue and red books is visualized in Fig. 1a. The monochromatic clustering reveals the political ‘echo chambers’ in a highly polarized population. (See Methods for methodological details, red/blue book comparisons that show similarity in publication year and sales rank, and information about the handful of ‘misfits’ in each cluster.)

The political and science books identified in amazon.com were also collected from barnesandnoble.com, to test the consistency of co-purchasing patterns between the two bookstores. These co-purchase networks, composed of millions of distinct purchases, are different in important ways: only 9% of Amazon co-purchase links are found in the Barnes & Noble network, and only 21% of Barnes & Noble links are found in Amazon. Nevertheless, the number of political links with each book in Amazon and Barnes & Noble is highly correlated (0.60). The consistency of our findings in these two environments suggests robust patterns of co-purchase behaviour, in which the political preferences of science book consumers are very similar even though the expression of those preferences in the purchase of particular books differs across the two websites.

We identified 428,433 titles that appeared under science categories in the Library of Congress and Dewey Decimal classifications. We grouped these into 27 exclusive high-level topics, corresponding to broadly defined scientific disciplines (such as physics, chemistry, medicine or economics). These 27 disciplines fall under four main scientific ‘schools’ (humanities, physical sciences, life sciences and social sciences). An additional 494,278 non-science titles were grouped in four main topics — arts, sports, literature (fiction and poetry), and religion — as a baseline for assessing co-purchase links between science and politics (see Methods for detailed categories).

We used co-purchase links to measure the political relevance, alignment and polarization among online customers. Political relevance of a topic is measured by the fraction of links from political books (whether red or blue) among all co-purchase links to books in the topic. Given the measure’s uncertainty with small samples, we used a Bayesian probabilistic framework with a prior distribution induced by the configuration model³². Links between science, non-science and political books are randomly generated, given the network degree of each book. The higher the political relevance, the

greater the likelihood that purchasers of political books will purchase books in the topic.

Political alignment of a topic is measured by the fraction of links from red books among co-purchase links from political books (red or blue). As with relevance, we account for statistical uncertainty by estimating this measure as the probability that books, in a particular topic will link with red books, conditioned on their links from political books. Alignment captures partisan interest in each scientific discipline on the red–blue spectrum, where purple (alignment = 0.5) could indicate the balanced political interest required for a public sphere of reasoned discourse.

Alternatively, a ‘purple’ discipline could also be internally polarized, with equal interest from left and right but in separate subsets of books. Political polarization is a function of the number of books within a discipline linked with both red and blue books, compared with a null model in which red and blue links are randomly assigned to books in the topic. Polarization equals 0 when disciplinary books are co-purchased with red and blue books uniformly at random, but increases as the sets of red- and blue-linked books diverge, indicating red and blue preferences for distinct books. (Note that we use the term ‘polarization’ to refer to the political segregation of the population into opposing groups of like-minded members who influence one another to adopt similar positions on other issues on which they did not initially agree³³. Our usage should not be confused with an alternative definition of polarization that indicates the tendency of people who agree on a given issue to influence one another to adopt more extreme opinions on that same issue.)

In addition to these three metrics, we also measured characteristics of scientific books and fields to account for diverging red and blue scientific interests. We scored fields as basic or applied science based on the ratio of patent to article citations in each field. We also considered whether books are ‘academic’ or ‘popular science’, based on publication by an academic or popular press. Finally, we measured the relative scientific breadth of liberals versus conservatives as the difference between the average number of disciplinary titles linked with a blue book and those with a red book, and the network location of books with ties to red and blue relative to the core or periphery of the co-purchase network within the disciplines. (See Methods for detailed discussion of these measures.)

Our analysis proceeds in three steps. We first assess political relevance, alignment and polarization as measures of political interest in science compared with political interest in books and

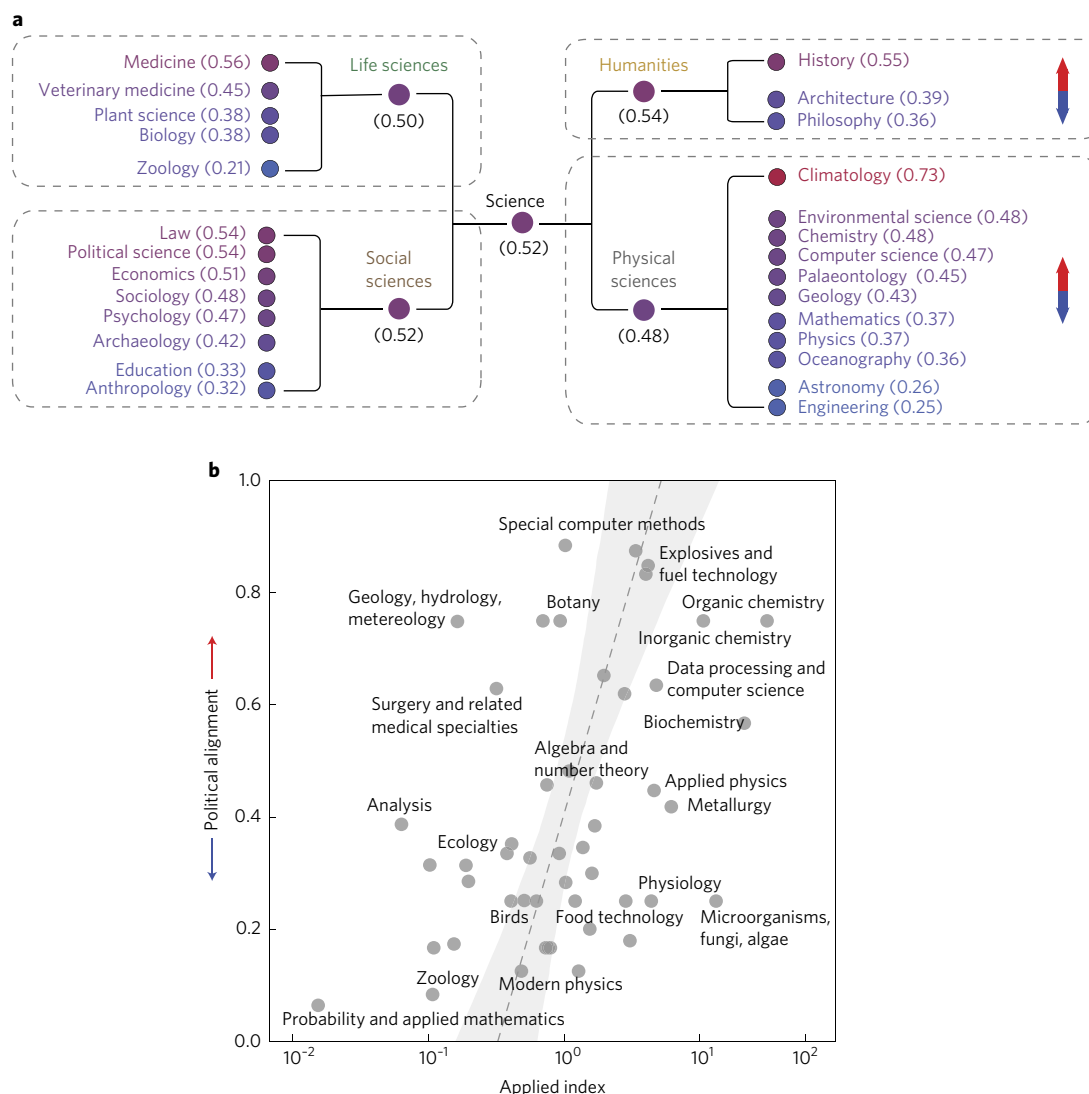


Figure 3 | Comparisons of political alignment across scientific disciplines and sub-disciplines. a, Colours and values in the disciplinary tree indicate political alignment. Differences in alignment emerge more clearly the more disaggregated the categories. Books in applied disciplines such as medicine, law and climatology are more likely to be co-purchased with conservative books, whereas those in basic science disciplines such as zoology, anthropology and philosophy are more likely to be co-purchased with liberal books. **b**, This association becomes clearer at the sub-discipline level, revealed in the scatter plot. Conservative alignment tends to increase with the applied index ($r = 0.43$, $p = 0.002$), which is defined as the ratio of patent to article citations, here shown on disaggregated Dewey Decimal subfield data along with a 90% confidence interval.

topics outside science; second, we report differences in these measures across scientific disciplines, broken down by the continuum of applied and basic science; third, we report the scientific breadth and location of red and blue books within disciplines.

First, compared with the number of co-purchase links expected by chance, people who buy liberal and conservative books are more likely to buy books on science than on important topics outside science, but this interest in science does not appear to be insulated from the ‘Big Sort’ of American politics. Figure 2 reports political relevance and polarization of science, religion, sports, arts and literature, with alignment indicated by colour. Results for political relevance show that political readers have greater interest in science relative to non-science topics, largely owing to books on social science. The physical and life sciences do not attract markedly greater interest among political readers compared with topics outside science, but this political interest in science is significantly more polarized than for arts and sports, indicating that liberals and conservatives are less likely to read the same science book.

Second, we found a significant positive correlation ($r = 0.43$, $p = 0.002$) between political alignment and the normalized number of citations by patents, our indicator of applied, commercial science (Fig. 3b). For example, organic chemistry is the most applied sub-discipline, as measured by patent versus article citations, and it aligns closely with red books (0.75 on an alignment scale from 0 to 1). This contrasts with a sub-discipline such as zoology, which is largely driven by curiosity and basic scientific concerns, and appeals more to those on the left (0.1). This pattern can also be observed in Fig. 3a, which reports co-purchase alignment within the four ‘schools’. Applied disciplines such as medicine and law attract readers at the red end compared with other disciplines in their respective schools, whereas anthropology and astronomy attract readers at the blue end. This mirrors the ideological differences among scientists employed in academia versus industry, as reported in AAAS/Pew surveys^{6,7} from 2009 and 2014. A possible interpretation is that scientific puzzles appeal more to the left, while problem-solving appeals more to the right.

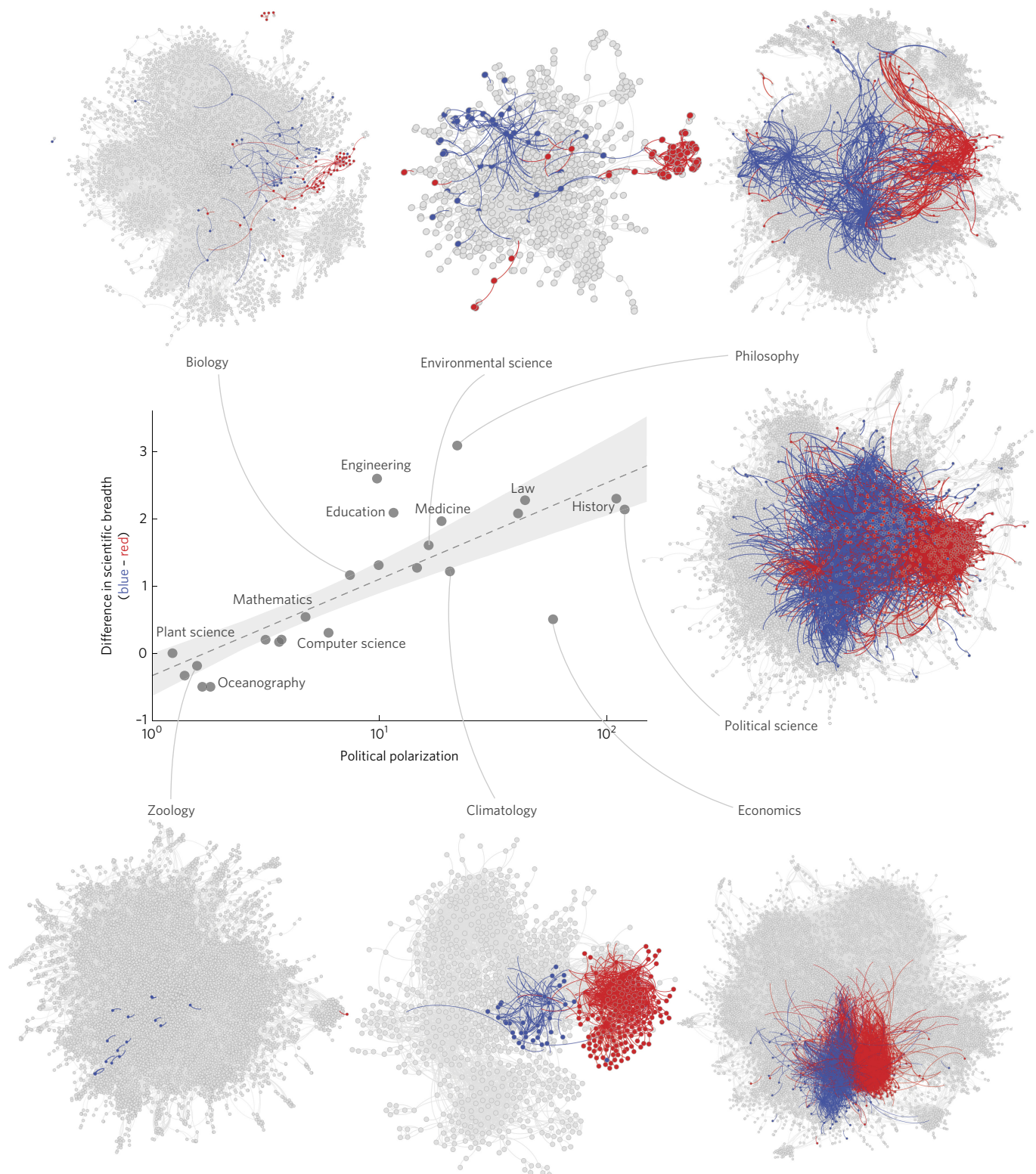


Figure 4 | Location of red and blue political books within the network of disciplinary books to which they are linked. Network visualizations show the locations of red and blue political books in the co-purchase networks of science books (grey nodes) in seven disciplines: zoology, biology, environmental science, climatology, economics, political science and philosophy. All disciplines show high polarization in partisan co-purchasing, with red books linked to fewer disciplinary books and to disciplinary books closer to the periphery (Supplementary Fig. 4). The peripheral location means that these red books are co-purchased with disciplinary books that are less likely to be co-purchased with other books in the discipline. The monochromatic clustering is due entirely to links among red-linked science books and likewise for blue-linked science books; the links between political books are not considered in constructing the networks. The scatter plot shows that the more polarized the political co-purchases, the higher the relative scientific breadth of blue versus red, as measured by the difference between the average number of disciplinary titles linked with a blue book and the number linked with a red book ($r = 0.8$, $p < 0.001$). The line represents the estimated relationship, along with the 90% confidence interval.

A few disciplines, notably palaeontology, bridge political divisions and are politically purple because books in these disciplines attract equal interest from both left and right. However, most purple disciplines (with equal likelihood to have red and blue book links) do not bridge political divisions. Figure 4 illustrates the internal network structure of seven natural and social science disciplines. Monochromatic clusters, located in different regions of the network, indicate that red and blue political books are linked to different interconnected clusters of disciplinary books. Simply put, even when left and right are equally likely to read books in a discipline, they are rarely the same books or even from the same topical cluster.

We drilled down further to examine the location of science books linked with red and blue books within co-purchase networks at the disciplinary level. The scatter plot in Fig. 4 presents the relative scientific breadth of blue versus red (y axis). Note that higher values in the y axis indicate greater difference between the breadths of blue and red books. The figure reveals that blue books link to a larger number of scientific titles than red within most disciplines, and the difference increases for more polarized disciplines.

The disciplinary networks show the location of red and blue political books linked to disciplinary books within a field (coloured grey). All books are positioned such that pairwise proximities correspond only to co-purchase links with disciplinary books. The visualizations show that red books tend to cluster on the periphery of the co-purchase networks for climatology, environmental science, political science and biology, which indicates that conservatives tend to purchase books likely to be co-purchased with each other but not with other books in the discipline (Supplementary Fig. 4). In contrast, blue books are less clustered and linked to books closer to the core, which indicates that liberals tend to purchase a more diverse set of science books, including books frequently co-purchased with other books in the discipline. The greater centrality of blue-linked books does not appear to be a consequence of academic liberalism. When books by academic publishers are removed, the pattern remains, as do all the other patterns that we report (see Supplementary Fig. 5).

These results need to be qualified by inherent limitations in the use of co-purchase links to measure partisan interest in science. First, although half of the US population purchases books online, this is not a random sample, which limits the ability to generalize our results to the other half. We cannot be certain that those who purchase books online are similar to those who do not, or that those who purchase red and blue books have the same interests in science as liberals and conservatives who do not purchase political books.

Second, we do not have individual-level co-purchase data and therefore cannot pursue analytical strategies for which such data would be required, such as multivariate causal modelling. Co-purchasing patterns reveal the population-level distribution of political interests in science but not the underlying individual-level causes.

Third, co-purchase links can influence purchases that then reinforce new co-purchase links, thereby magnifying estimates of the underlying consumer interest in politics and science. We believe that this reinforcing dynamic does not invalidate the patterns that we find. Rather, this provides a possible explanation of the findings — an ‘echo chamber’ constructed not only by consumer preferences but also by the ‘filter bubble’ constructed by collaborative filtering technology that reinforces those preferences³⁴. Future research using randomized trials is needed to test whether recommender systems can induce users to make choices orthogonal to or counter to their existing preferences.

These three sets of concerns are mitigated, however, by similarity in the results derived from distinct bookstores, with different purchase patterns and company-specific algorithms. We replicated the Amazon-based analyses using the Barnes & Noble network and

found consistent results across datasets, with political polarization ($r = 0.97$, $p < 0.001$) and alignment ($r = 0.76$, $p < 0.001$) highly correlated across the two websites (Supplementary Fig. 6).

Keeping the limitations of these data in mind, online co-purchases reveal a similar level of interest in science among liberal and conservative readers, but not in the same topics. We found little support for the view that science is either an apolitical separate sphere, largely ignored by partisans, or a public sphere in which left and right share common scientific interests. Books on science are more likely than novels or books on religion, sports or the arts to be co-purchased with political books, but left and right rarely purchase the same books. Scientific preferences are polarized at the aggregate level, with liberals attracted to basic science, and conservatives attracted to applied, commercial science. Books in a few ‘purple’ disciplines such as palaeontology are co-purchased by both the left and right, but these ‘bridging’ disciplines also tend to be those with the lowest relevance to political readers. Disciplines with high political relevance, such as social science, law, history and biology, tend to attract politically aligned readers, and even when they attract readers from both left and right, it is not to the same books. Although our analysis cannot uncover the cognitive and social processes underlying these partisan differences in the consumption of science, it is consistent with research on ideologically motivated reasoning³⁵, in which political differences seed different predilections toward politically consequential scientific positions or amplify path-dependent routes of political and cultural sorting¹⁹.

Our finding that red books tend to link to a narrow cluster of relatively peripheral books in the more polarized disciplines is consistent with recent research documenting growing scepticism among conservatives of the cultural authority held by professional scientists¹⁶. This may also reflect efforts by conservative political movements to offer politicized alternatives to consensus scientific positions. Such efforts have been observed in previous research on climate change¹⁰ and a range of other areas³⁶. Science may not be on the front lines of the culture wars, but it is not above the battle, nor is it immune to the ‘echo chambers’ that have been widely observed in political discourse.

This conclusion underscores the need for research into possible remedies. Scholars have begun to explore counter-measures of scientific communication that include helping scientists to establish shared interests with audiences to enhance credibility³⁷, adding public deliberation alongside scientific analysis to separately identify fact and value^{38,39}, and communicating consensus when it exists to help the public set aside protective motivations¹⁴. It is hoped that these will counter selective exposure to ‘convenient truth’ and renew the promise of science to inform and elevate political debate.

Methods

Data collection. We collected metadata for 1,449,525 books from the largest online book retailer, Amazon.com, in spring 2013. Starting from two seed books, one liberal (Barack Obama’s *Dreams from My Father*) and one conservative (Mitt Romney’s *No Apology*), we collected all information accessible from the webpage of each book, including descriptive information, reviews, and the list of up to 100 books under “Customers Who Bought This Item Also Bought”. We then collected identical data for each book on the recommended list, iterating that process until no new books could be identified. The resulting ‘snowball sample’ contained virtually all books in the largest strongly connected component in Amazon’s directed co-purchase network.

Because every title may have multiple editions or formats, for example paperback and hardcover, each of which is associated with a distinct ISBN (International Standard Book Number), we consolidated different editions and formats based on the unique ASIN (Amazon Standard Identification Number) provided in the source code of each book page. After consolidation we ended up with 1,303,504 unique titles in the dataset. In the main text and below, we use ‘books’ to refer to the distinct titles and never to the physical objects.

The political and science books identified in amazon.com were also collected from barnesandnoble.com, together with books from the Barnes & Noble

co-purchase recommendation lists for those books. See below for how science and political books are identified. A brief summary of the two datasets is given in Supplementary Table 1.

Political books. Our 3,714 candidate political books were chosen from three sources: 1,667 books listed in Amazon's 'Liberalism & Conservatism' category; 1,812 additional books listed in 20 Amazon categories that share the most books with 'Liberalism & Conservatism'; and 320 books written by prominent US political leaders, including members of the US Congress since 1993 and major party presidential candidates since 1992.

Each candidate political book was examined by at least two independent coders, with another independent coder as a tiebreaker. The coders classified each book as liberal, conservative or ideologically indeterminate, based on the information available on the book's Amazon webpage. For a book to be coded as liberal or conservative, it had to meet two basic requirements. First, the book must have political content. That is, the book must express an ideological position on a partisan political or social issue, even if that issue is not necessarily the main topic of the book. For example, autobiographies are nominally about a person, but autobiographies of political figures often express ideological positions on contentious political issues, in which case the book meets the requirement to have political content. Second, the ideology that the book espouses must be consistently liberal or conservative. If the book expressed contradictory ideological positions, or centrist positions that are shared by both liberals and conservatives or by neither, the book was coded as indeterminate.

Two independent coders reached agreement on 83.5% of the candidate books listed in Amazon's 'Liberalism & Conservatism' category. Another two independent coders reached agreement on 70% of the books in the 20 categories that overlap with 'Liberalism & Conservatism'. The lower agreement reflects the greater diversity among books that were not listed by Amazon under 'Liberalism & Conservatism'. Conflicts were resolved by other independent coders and, in rare cases, by the authors. In total, we identified 677 conservative ('red') books, 587 liberal ('blue') books, and 2,545 indeterminate books. Details of the conservative and liberal books are given in Supplementary Tables 3 and 4.

We tested for possible bias introduced by the exclusion of political books that did not meet both our criteria for red–blue classification, by including 315 'indeterminate' books written by US Congressmen coded red or blue based on the authors' party memberships. We found that adding these 'author-coded' books to the 'content-coded' books did not weaken the reported results (see Supplementary Fig. 8).

Distributions of sales ranks for liberal and conservative books are shown in Supplementary Fig. 2, along with distributions of publication years for the two sets of books. The average logarithmic sales ranks for liberal and conservative books are 13.5 and 13.1, respectively, and the medians are 13.7 and 13.4. The average publication years of both sets of books are 1999, and the median publication year is 2007 for liberal books and 2009 for conservative books. A summary is provided in Supplementary Table 2.

Science books and categories. We consolidated the science-related categories in the Library of Congress and Dewey Decimal classification systems and reorganized these categories into 27 exclusive high-level topics, corresponding to broadly defined scientific disciplines (such as physics, chemistry, medicine or economics). These 27 disciplines fall under four main scientific 'schools' (humanities, physical sciences, life sciences and social sciences). We then used the Library of Congress and Dewey Decimal codes of the books to sort them into disciplines.

An additional 494,278 non-science titles were grouped in four main topics — arts, sports, literature (fiction and poetry) and religion — as a baseline for assessing co-purchase links between science and politics. (Detailed organization of the Library of Congress and Dewey Decimal categories can be found in Supplementary Tables 5 and 6.)

Political relevance. Political relevance measures how likely it is that books from a given topic will be co-purchased with political books. It is a number between 0 and 1, defined as the probability θ that a co-purchase link from books in a given topic will link that topic with political books (red or blue). More formally, we assume that the number of undirected co-purchase links X between the topic (for example climatology) and political books has a binomial distribution, $X \sim \text{binomial}(K, \theta)$, where K is the total number of links attached to the topic and \sim indicates 'distributed as'.

A straightforward estimator of this probability θ is the number of co-purchase links between the topic and political books divided by the total number of co-purchase links between the topic and all other topics, X/K . However, this estimator is not appropriate in this application for two reasons. First, a topic might have few links to other topics, which renders this measure of relevance unreliable. For example, if a topic has only one link to other topics and it links to political books, then we have $\theta = X/K = 1$, implying that the topic is extremely relevant, which is dubious because there is only one co-purchase link in total. Second, although the uncertainty of this estimator resulting from too few observations could be captured by a confidence interval, we needed to compare relevance

across topics while still taking this uncertainty into consideration. We needed to normalize the scores across topics. For these two reasons, we developed a Bayesian estimate for the probability θ . The procedure is as follows. First, we assume a null model in which all the co-purchase links between topics are randomly shuffled while preserving the number of links attached to each topic (compare with the configuration model³²). This 'best guess' of the co-purchase network (without knowing the identities of the topics) represents a prior distribution on θ , $\theta \sim \text{beta}(dk_{\text{political}}/m, d(1 - k_{\text{political}}/m))$, where $k_{\text{political}}$ is the total number of links attached to political books and m is the total number of links between topics (including political books). $k_{\text{political}}/m$ is the probability of linking to political books in our random null model, and d controls the strength of our prior. The values of θ will depend on d , but we are interested in the relative size of θ across topics, and thus our results are insensitive to the choice of d . We use the average number of links to political books over all disciplines as d . After observing the number of links between the topic and political books, we can update our estimate of θ using Bayes rule:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

We can then derive a Bayesian estimate for the probability θ :

$$\theta \sim \text{beta}(dk_{\text{political}}/m, d(1 - k_{\text{political}}/m))$$

$$X|\theta \sim \text{binomial}(K, \theta)$$

$$\theta|X \sim \text{beta}(X + dk_{\text{political}}/m, K - X + d(1 - k_{\text{political}}/m))$$

where the posterior distribution $\theta|X$ combines our initial guess of θ (based on the randomized network) with the actual data X and weights the actual data by the number of co-purchase links attached to the topic.

Finally, the political relevance of the topic is defined as the mean of the posterior distribution of θ :

$$E[\theta|X] = \frac{X + dk_{\text{political}}/m}{K + d}$$

This model further reflects how the actual observations (X/K) are combined with prior beliefs ($k_{\text{political}}/m$) to incorporate uncertainty resulting from too few observations on the point estimator of θ . If the number K of co-purchase links attached to a topic is small, we do not have enough evidence to estimate its political relevance, and this will be close to the level in the random null model; if a topic has many co-purchase links, we have greater trust in the estimate.

Political alignment. Political alignment locates each discipline on the blue–red spectrum based on the interest among conservative readers relative to liberal. Analogous to political relevance, alignment is defined as the probability that a co-purchase link attached to a book in a given topic is to a red book, conditioned on the link being to a political book (red or blue). Note that we restrict our focus to links with political books rather than to all topics. Hence, it can be viewed as a measure of how relevant the topic is to readers of red books as opposed to blue books. For ease of notation, we also denote this probability θ . One way to estimate this probability θ is to divide the number of links (X_{red}) between the topic and red books by the total number of links (K_p) between this topic and political books, X_{red}/K_p . Similar issues arise as for political relevance, however, and hence we developed an analogous Bayesian model to estimate θ .

First, we assume a null model in which co-purchase links between a given topic and political books are randomly shuffled while preserving the number of links attached to each (compare with the configuration model³²). Let k_{red} and k_{blue} be the total numbers of links attached to red and blue books, respectively. Then the probability of linking to red books in the random null model is $k_{\text{red}}/(k_{\text{red}} + k_{\text{blue}})$, which suggests a prior distribution on θ ,

$$\theta \sim \text{beta}(dk_{\text{red}}/(k_{\text{red}} + k_{\text{blue}}), dk_{\text{blue}}/(k_{\text{red}} + k_{\text{blue}}))$$

where d controls the strength of our prior assumption. After observing the number of links (X_{red}) between the topic and red books, and the number of links (X_{blue}) between the topic and blue books, we update our knowledge on θ and obtain the posterior distribution of θ . In summary, the Bayesian model for estimating θ is as follows:

$$\theta \sim \text{beta}(dk_{\text{red}}/(k_{\text{red}} + k_{\text{blue}}), dk_{\text{blue}}/(k_{\text{red}} + k_{\text{blue}}))$$

$$X_{\text{red}}|\theta \sim \text{binomial}(K_p, \theta)$$

$$\theta|X_{\text{red}} \sim \text{beta}(X_{\text{red}} + dk_{\text{red}}/(k_{\text{red}} + k_{\text{blue}}), K_p - X_{\text{red}} + dk_{\text{blue}}/(k_{\text{red}} + k_{\text{blue}}))$$

Accordingly, political alignment is calculated as the mean of the posterior distribution:

$$E[\theta|X_{\text{red}}] = \frac{X_{\text{red}} + dk_{\text{red}} / (k_{\text{red}} + k_{\text{blue}})}{K_p + d}$$

We note one extra step in presenting the alignment score. Naively, one might expect a topic to be right-leaning if the topic has a larger probability of linking with red books than blue books (that is, $\theta > 0.5$), but this is not a fair comparison. Because red books have more total co-purchase links than blue books, red books would also possess more co-purchase links in the random null network. Therefore, the fair comparison is to compare θ with $k_{\text{red}} / (k_{\text{red}} + k_{\text{blue}})$, which is the probability of linking to red books in the random null model. To make the presentation of results more accessible and intuitive, we scaled θ linearly so that when $\theta = k_{\text{red}} / (k_{\text{red}} + k_{\text{blue}})$ the rescaled θ would be 0.5. In the main text and in the Supplementary Information, θ is always rescaled and 0.5 is the 'neutral' point, which accords with intuition. The rescaling is straightforward: θ is rescaled to $0.5\theta / [k_{\text{red}} / (k_{\text{red}} + k_{\text{blue}})]$ when $\theta < (k_{\text{red}} + k_{\text{blue}})$, and to $0.5[\theta - k_{\text{red}} / (k_{\text{red}} + k_{\text{blue}})] / [1 - k_{\text{red}} / (k_{\text{red}} + k_{\text{blue}})] + 0.5$ otherwise.

Lastly, we checked the consistency of our measurements of political alignment by measuring the alignment of all political books and comparing the results with the 'ground truth' red and blue hand-codings. Specifically, for every political book that had been hand-coded as blue or red, we ignored the 'ground truth' and instead imputed the book's alignment using the same procedure for measuring alignment of scientific topics introduced above. We then classified a book with alignment $\theta > 0.5$ as red and otherwise blue. The imputed ideology agreed with human codings for over 96% of the political books. Inspection of the handful of anomalies revealed red books that appeal to liberals (for example conservative criticisms of religious political influence) and blue books that appeal to the Tea Party (for example Alinsky-inspired community organizing).

Political polarization. Political polarization measures the extent to which interests from conservatives and liberals in a given topic diverge. In other words, even if conservatives and liberals are equally likely to buy books of a given topic, they might buy distinct, ideology-relevant books. Polarization identifies this possibility. For a given topic, we computed the number of books within the topic linked to both red and blue books, divided by the number of books linked to either (we call this quantity 'overlap', denoted by o). We then compared observed overlap with the expected value in a null model in which links from red and blue books are randomly assigned to books in the topic.

Specifically, for each link between books in a given topic and political books, we shuffle the book in the topic to a randomly chosen topical book, also linked to political books. This results in all politically relevant books from the topic being linked to political books uniformly at random. After all political links have been randomized, the fraction of books linked to both red and blue among books linked to either is calculated in the randomized network. We carry out 100 such random simulations to obtain a distribution of the overlap between blue and red (the fraction of books linked to both blue and red) in the random model. Finally, polarization is measured as the difference between the expected overlap in the random null model, $E[o]$, and the observed overlap:

$$\frac{E[o] - o}{\sqrt{\text{var}[o]}}$$

If the observed overlap is smaller than what would be expected at random, polarization is positive and increases with the difference between expected and observed overlaps. Polarization of a topic equals zero when red and blue books are co-purchased with books within the topic uniformly at random, but increases as the sets of red- and blue-linked books diverge, indicating red and blue preferences for distinct books.

Applied index. To quantify the extent to which the political alignment of scientific fields is correlated with their application, we developed an index measuring the extent to which a field is commercially applied.

Among many possible measures of commercial application, a reasonable and tractable one is the degree to which patents build upon knowledge produced by the field. To that end, we use the US patent database from Google, 1976–2014, which is the most digitally complete. For each journal, we tabulate the number of times that journal was cited by all patents in the patent database. We aggregated these within the Dewey Decimal and Library of Congress categories in which Amazon books are categorized, such that the number of citations received by a field is computed as the sum of citations received by all the journals in that field.

Citations from patents meaningfully reflect gross contributions of commercial relevance to the real world, but they do not fully capture the degree to which a field is applied, because the number of citations is strongly influenced by field size and activity. Imagine a small, focused field that is commercially applied, compared with a large, broad field. Perhaps the smaller, more focused field receives fewer citations from patents only because it produces fewer total patents, but it is actually more application-oriented because all knowledge in the field is

transferred to technology. Therefore, we build a field-level citation network to capture how active or impactful a field is in the scientific space.

We find that the number of citations that a sub-discipline (as defined by our classification of science categories) receives from patents is strongly correlated ($r = 0.8$, $p < 0.001$) with the number of citations that it receives from other sub-disciplines (Supplementary Fig. 3). This correlation reveals that most sub-disciplines are proportionally active in both patent and academic domains, and either of the two types of citations alone is not sufficient to estimate how applied the field might be. Accordingly, we constructed a commercial applied index that combines both kinds of citations and measures how much each sub-discipline is cited by patents relative to articles. Specifically, for each sub-discipline i , we denote the number of citations from patents by y_i and the number of citations from articles in other sub-disciplines by x_i . The expected number $E[Y]$ of citations from patents given the number X of citations from other sub-disciplines is modelled $E[Y] = \exp(aX + b)$, which explains the correlation illustrated in Supplementary Fig. 3. Finally, the commercial application index A of a sub-discipline is calculated as the number of citations from patents normalized by its expected number of patent citations given the number of citations from other sub-disciplines: $A_i = y_i / \exp(ax_i + b)$.

Scientific breadth. Scientific breadth measures the breadth of interests in science from conservatives and liberals. For example, if red books link to a narrow subset of books within a discipline, while an equal number of blue books connect to a large and diverse subset of disciplinary books, those purchasing blue books have exposure to a wider range of science books — and probably a wider range of scientific perspectives — than those purchasing red books.

With respect to each discipline, the scientific breadth of conservatives (or liberals) is defined by the number of distinct titles within the discipline linked to red (or blue) books divided by the number of red (or blue) books linked to the discipline.

Centrality. The network location of each book with links to red and blue relative to the core or periphery of a disciplinary book network was quantified by assessing the closeness centrality⁴¹ of each book with respect to the given disciplinary network. Centralities of blue-linked books were then compared with those of red-linked books within each discipline to assess core location of blue and red with respect to the discipline. Books at the centre of a disciplinary book network can be interpreted as having greater disciplinary relevance, given their co-presence in personal libraries with many other books in the discipline. In contrast, books at the periphery are rarely purchased with other books in the discipline.

To compare the difference between centralities of blue- and red-linked books across disciplines, we calculated the widely used t -score

$$\frac{E[X] - E[Y]}{\sqrt{S^2(1/n_x + 1/n_y)}}$$

for each discipline, where X corresponds to centralities of blue-linked books, Y to centralities of red-linked books, n_x to the number of blue-linked books and n_y to the number of red-linked books; S^2 is the pooled sample variance of blue- and red-linked books.

The standardized difference (t -score) between centralities of blue- and red-linked books is shown in Supplementary Fig. 4. The mean centrality of blue-linked books is larger than that of red-linked books for most disciplines, and as polarization increases, blue-linked books are more central than red-linked books. Exceptions to this pattern include history and economics.

Data availability. All data that support the findings of this study are publicly available at http://www.knowledgelab.org/research_grants/research/amazon/

Code availability. Python code for data processing, analysis and visualization is publicly available at http://www.knowledgelab.org/research_grants/research/amazon/

Received 1 September 2016; accepted 21 February 2017; published 3 April 2017

References

1. Daston, L. & Galison, P. *Objectivity* (Zone Books, 2007).
2. Habermas, J. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (MIT Press, 1991).
3. Shapin, S. A *Social History of Truth: Civility and Science in Seventeenth-Century England* (Univ. Chicago Press, 1994).
4. Shapiro, B. J. A *Culture of Fact: England, 1550–1720* (Cornell Univ. Press, 2003).
5. Sutton, G. V. *Science for a Polite Society: Gender, Culture, and the Demonstration Of Enlightenment* (Westview, 1997).
6. Kohut, A., Keeter, S., Doherty, C., Dimock, M. & Leshner, A. I. A *Survey Conducted in Collaboration with the American Association for the Advancement of Science* (Pew Research Center for the People & the Press, AAAS, 2009).

7. Funk, C. *et al.* *Public and Scientists' Views on Science and Society* (Pew Research Center for the People & the Press, AAAS, 2015).
8. *Public Praises Science; Scientists Fault Public, Media* (Pew Research Center for the People & the Press, AAAS, 2009).
9. Fiorina, M. P. & Abrams, S. J. Political polarization in the American public. *Annu. Rev. Polit. Sci.* **11**, 563–588 (2008).
10. Farrell, J. Corporate funding and ideological polarization about climate change. *Proc. Natl Acad. Sci. USA* **113**, 92–97 (2015).
11. Suhay, E. & Druckman, J. N. The politics of science political values and the production, communication, and reception of scientific knowledge. *Ann. Am. Acad. Pol. Soc. Sci.* **658**, 6–15 (2015).
12. Oreskes, N. & Conway, E. M. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming* (Bloomsbury USA, 2011).
13. Jasanoff, S. Technologies of humility. *Nature* **450**, 33 (2007).
14. Bolsen, T. & Druckman, J. N. Counteracting the politicization of science. *J. Commun.* **65**, 745–769 (2015).
15. Berezow, A. B. & Campbell, H. *Science Left Behind: Feel-Good Fallacies and the Rise of the Anti-Scientific Left* (PublicAffairs, 2012).
16. Gauchat, G. Politicization of science in the public sphere: a study of public trust in the United States, 1974 to 2010. *Am. Sociol. Rev.* **77**, 167–187 (2012).
17. Bishop, B. *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart* (Mariner, 2009).
18. Bakshy, E., Messing, S. & Adamic, L. Exposure to ideologically diverse news and opinion on Facebook. *Science* **384**, 1130–1132 (2015).
19. DellaPosta, D., Shi, Y. & Macy, M. Why do liberals drink lattes? *Am. J. Sociol.* **120**, 1473–1511 (2015).
20. Yeo, S. K., Xenos, M. A., Brossard, D. & Scheufele, D. A. Selecting our own science: how communication contexts and individual traits shape information seeking. *Ann. Am. Acad. Pol. Soc. Sci.* **658**, 172–191 (2015).
21. Jelveh, Z., Kogut, B. & Naidu, S. Political language in economics. Preprint at <http://dx.doi.org/10.2139/ssrn.2535453> (2014)
22. Massey, D. S. & Tourangeau, R. Where do we go from here? Nonresponse and social measurement. *Ann. Am. Acad. Pol. Soc. Sci.* **645**, 222–236 (2013).
23. Vanraan, A. F. J. Fractal dimension of co-citations. *Nature* **347**, 626–626 (1990).
24. Jenssen, T.-K., Lægreid, A., Komorowski, J. & Hovig, E. A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.* **28**, 21–28 (2001).
25. Chen, C. Searching for intellectual turning points: progressive knowledge domain visualization. *Proc. Natl Acad. Sci. USA* **101**, 5303–5310 (2004).
26. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
27. Sinatra, R., Deville, P., Szell, M., Wang, D. & Barabási, A.-L. A century of physics. *Nat. Phys.* **11**, 791–796 (2015).
28. Eakin, E. Study finds a nation of polarized readers. *The New York Times* (13 March 2004).
29. Krebs, V. *The Social Life of Books: Visualizing Communities of Interest via Purchase Patterns on the WWW* (OrgNet, 1999); <http://www.orgnet.com/booknet.html>
30. Krebs, V. *Divided We Stand?* (OrgNet, 2003); <http://www.orgnet.com/leftright.html>
31. Linden, G., Smith, B. & York, J. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Comput.* **7**, 76–80 (2003).
32. Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **6**, 161–180 (1995).
33. Sunstein, C. R. *Republic.com 2.0* (Princeton Univ. Press, 2009).
34. Pariser, E. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (Penguin, 2011).
35. Kahan, D. M. Ideology, motivated reasoning, and cognitive reflection: an experimental study. *Judgm. Decis. Mak.* **8**, 407–24 (2012).
36. Jacques, P. J., Dunlap, R. E. & Freeman, M. The organisation of denial: conservative think tanks and environmental scepticism. *Environ. Polit.* **17**, 349–385 (2008).
37. Lupia, A. Communicating science in politicized environments. *Proc. Natl Acad. Sci. USA* **110**, 14048–14054 (2013).
38. Dietz, T. Bringing values and deliberation to science communication. *Proc. Natl Acad. Sci. USA* **110**, 14081–14087 (2013).
39. Maibach, E. W. & van der Linden, S. L. The importance of assessing and communicating scientific consensus. *Environ. Res. Lett.* **11**, 091003 (2016).
40. Alinsky, S. *Rules for Radicals* (Knopf Doubleday, 2010).
41. Sabidussi, G. The centrality index of a graph. *Psychometrika* **31**, 581–603 (1966).

Acknowledgements

We are grateful for comments from participants in seminars at Microsoft Research–New England, MSR-NY, GESIS-Koln, University of Michigan School of Information, Duke University DNAC, the first International Conference on Computational Social Science, and the Computational Social Science Summit at Northwestern. We acknowledge funding from the John Templeton Foundation to the Metaknowledge Network, NSF SES 1303533, SES 1226483, SES 1158803, National Research Foundation of Korea NRF-2013S1A3A2055285 and Air Force Office of Scientific Research FA9550-15-1-0162, and computation support from the Open Science Data Cloud. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

All authors designed the research and drafted the paper. E.S., Y.S. and F.A.D. gathered and analysed the data.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.A.E and M.W.M.

How to cite this article: Shi, F., Shi, Y., Dokshin, F. A., Evans, J. A. & Macy, M. W. Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nat. Hum. Behav.* **1**, 0079 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Competing interests

The authors declare no competing interests.