Peer Review File

Manuscript Title: Papers and patents are becoming less disruptive over time

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

Using large-scale datasets, the authors investigated whether progress in science is slowing in major fields or not. The analyses are mainly based on the newly introduced disruption indicator. The study seems technically and methodologically sound. However, I have some suggestions for improvements:

Lines 61-62: It is not clear what the percentages mean. Why didn't the authors present mean CD index values instead? Furthermore, I am not sure whether CD index values can be compared across decades. The publication and citation cultures (and database coverages) are so different that these comparisons might be impossible. In bibliometrics, it is conventional to time-normalize indicators – besides field-normalization (Waltman, 2016). I propose to do the same with disruption indicators (see Bornmann & Tekles, 2019).

Lines 69-71: This analysis could be connected to the incommensurability hypothesis of Thomas Kuhn (Wray, 2011). Further indicators that could be used for robustness checks in the study are novelty indicators (Wang, Veugelers, & Stephan, 2017). The development of these indicators has a long tradition in scientometrics, and the indicators are conceptually close to disruption.

Lines 84-85: In my opinion, "produce" is not a verb with a (close) link to disruption. I cannot follow the rational of using this word as a link to disruption. The same is true for the word "use" and its link to consolidating.

Figure 2: I see sharp increases of CD values in 1960 for social sciences and around 1980 for technology. Please explain these increases.

Lines 216-217: This paper has been published (Leydesdorff, Tekles, & Bornmann, 2021).

Lines 304-305: What is the reason for the random selection?

Lines 420-422: A paper with more recent results exist (Bornmann, Haunschild, & Mutz, 2021).

Page 26: This is wrong: "For papers, lines correspond to WoSresearch areas; for patents, lines correspond to NBER technology categories".

Page 28: Figure B is especially interesting. Until 1960, mean DI values are (significantly) higher than

0. Since then, they are around zero. Furthermore, papers with DI = 1 are especially visible before 1960. I wonder how these results can be explained. Is there a change in the structure of citation data?

Page 30: Several CD index variants are discussed in both papers. Please explain which have been used in the study. Give reasons why you preferred certain variants.

Section 3: Here, it is not clear what the authors did. Why did they involve two researchers who independently searched for words? Furthermore, the survey is not clear. How many people have been asked, and answered? Who has been asked? Researchers? What kind of sample has been surveyed?

References

Bornmann, L., Haunschild, R., & Mutz, R. (2021). Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. Humanities and Social Sciences Communications, 8(1), 224. doi: 10.1057/s41599-021-00903-w.

Bornmann, L., & Tekles, A. (2019). Disruption index depends on length of citation window. El profesional de la información, 28(2), e280207. doi: 10.3145/epi.2019.mar.07.

Leydesdorff, L., Tekles, A., & Bornmann, L. (2021). A proposal to revise the disruption indicator. Profesional de la Información, 30(1). doi: 10.3145/epi.2021.ene.21.

Waltman, L. (2016). A review of the literature on citation impact indicators. Journal of Informetrics, 10(2), 365-391.

Wang, J., Veugelers, R., & Stephan, P. E. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Research Policy, 46(8), 1416-1436.

Wray, K. B. (2011). Kuhn's evolutionary social epistemology. Cambridge, UK: Cambridge University Press.

Referee #2 (Remarks to the Author):

Review of “the decline of disruptive science and technology” by Park, Leahey and Funk for Nature.

In “the decline of disruptive science and technology” the authors use multiple measures, including an interesting one that they recently invented (and has received widespread attention), to trace a declining trend in science and technology over the latter half of the 20th Century. They analyze this in the context of citation and word-use patterns across multiple datasets, attempt to control for confounding explanations, and try to identify explanations, painting a coherent portrait of science and technology institutions that have slowed as they have grown, aged and narrowed.

There are several things to like about the manuscript, including the careful, redundant measurement across fields (and even domains, e.g., JSTOR vs. the Web of Science us. USPTO). Nevertheless, I had some concerns about the manuscript and its claims with respect to measurement, effect identification, interpretation and framing.

On measurement, the Consolidating/Disrupting (CD) measure, which forms the core of the author's empirical argument is based on a quantity that necessarily decreases with the amount of science available to be cited (the "b" of prior works cited--a variable stated in Figure 1, but not defined across the paper). In other words, while their interpretation of the measure as an indication that a word is viewed as representing a new direction in science or technology is valid at any given point of time (e.g., for two focal pieces at any given time), its comparison across time is suspect because the amount of science at risk of being cited, and relevant to current science is increasing over time, which can only trivially be interpreted as a reduction in innovation. In other words, with more work being cited earlier because there is more relevant work to be cited (and an aging scientific workforce that is citing older work), the disruption measure feels less like an indicator of new views in science emerging over time, than a shift in the measurement of scientific views. In other words, there are more (citational) stories to tell about how recent work is tied in with the past over time. The authors take pains to reduce the historical variation in the measure by only looking at the citations 5-years out, but they cannot as easily account for the changing relevance of the past. Incidentally, the same historical pattern is demonstrated in Extended Data Fig. 3 of (Wu, Wang, and Evans 2019) in each panel c through m. The authors state that "these declines demonstrate that relative to earlier eras, recent papers and patents do less to push science and technology in new directions". This feels more directly demonstrative that there are more possible stories of progress—where the present builds on the past—to tell about any particular recent paper. A further measurement issue was the lack of a clear definition of their CD index (I had to go back to their prior 2016 paper to make sense of the variables used in Figure 1 and the calculation it describes), and the failure to describe and illustrate the distribution of this measure at any one point in time (other than tracing its mean in Figure 2), the distribution of varied measures to show their correspondence (e.g., CD at other times that 5), and their stating (erroneously) that the median value is 0 in Figure 1.

To be clear, I liked the word diversity measure (types/tokens) and its consistent pattern over time, which provides a correspondence with the CD results, but the effects become much more modest, especially after 1970 for papers, and 1990 for patents (Figure 3), which are not described within the paper and have an interesting, but undescribed or theorized punctuated form (with a "kink" at 1970).

The second analysis of verb classification and scoring was less persuasive to this reviewer. In fact, it was not even totally obvious whether this just one or two separate analyses (described in the section half of SI Section 2 and Section 3). The selection of words seemed idiosyncratic and complex in order to "eliminiate idiosyncratic terms" (SI:506). The procedure involved including verbs (other parts of speech not presented because "verbs…generally yielded more substantively interesting patterns") that significantly correlated with utilization over time, occurred in more than 1000 papers, and excluded “topical terms” that traced the changing “substance” of science rather than its “approach”. Then they performed a survey with Science of Science researchers they say is described in Methods (didn’t find there, but I did in the next section of the SI…but wasn’t sure if that was the same or a different analysis). The survey they described in the next section was performed on the All Our Ideas platform, which fields a “wiki-survey” that performans an pairwise ordinal comparison task (is verb A “more indicative of efforts to create knowledge/technology” than B), sequentially repeated with light (Bayesian) active learning and designed for the crowd-introduction of new topics for comparison. The concern with its use here is that the population of those surveyed was “three researchers who are familiar with the literature on technology strategy and Science of Science.” I assume that these are not the three authors on this paper, also presumably familiar with the literature on technology strategy and the Science of Science. This wouldn’t necessarily be a problem because the authors chose the 2 sets of categories (about “creation”, “discovery” and “perception” for disruption, and “assessment”, “improvement”, and “application” for consolidation), but these are not “validated” by the three researchers, but rather the (tiny) crowd is simply used to source words within each category that are most related to “efforts to create knowledge/technology”. There is no indication of how many repetitions the survey was performed; how many unique pairs were floated, or the sequence with which elements were added by the researchers (presumably none were added by the participants, despite this being the primary justification for the active learning approach in All Our Ideas…the survey was designed to converge slowly so that new things could be added and have a chance of ascendancy in the overall score even if added late; there are faster-converging methods when no addition is involved, but we don’t know if fast convergence was required because the number of surveyed pairs is not mentioned). Then the top 25 words with the highest score in each disruptive category “were all deemed more likely to appear in disruptive titles” and those highest in each consolidating category “were all deemed more likely to appear in consolidating titles”. I’m not sure how this was assessed; maybe from Table 1 (but this focused on the increasing and decreasing terms?) or what statistics were used, or whether or how significantly the associations were. Incidentally, it would have been more persuasive to this reviewer if the survey was used to validate the author’s priors about words and word categories and not simply help them characterize their prior. Moreover, the words themselves did not appear semantically distinct or self-describing disruptive innovation vs. consolidation. Take the words in Supp Table 1…its not obvious that “pour”, “label”, and “note” are more or less disruptive than “sing” “surface” and “disseminate”? “Produce” is the main example from the paper as a disruptive word, but it feels polysemous—are we producing something new for the very first time (but then would we be using the word produce rather than invent or discover?) or are we producing something long-discovered at scale, in which case it becomes conceptually more similar to “automate”. In short, (for me) the second semantic analysis created more heat than light.

The authors then identify historical decreases in the diversity (normalized entropy) of work cited, increases in the number of self-citations, and increases in the age of cited work. They regress disruption on diversity, self-citation and aging work and find that more diverse, new-to-the-authors, and recent work anticipate disruption (at 5 years) despite changes moving in the opposit direction. I feel that the diversity/disruption findings is an important contribution and demonstration, and it would have been nice for this path-dependant change in the entropy to have been better described (is everyone citing the same things, and are those things semantically like one another?) This was the big take-away for me from the paper. The other findings feels someone oddly interpreted: is self-citation an indication that scientists and inventors are more focused on their own work, or there is more of that work to be cited (as scientists age). Moreover, are aging citations an indication that scientists “struggle to keep up” or that the stock of relevant work is increasing (see my first point above) and/or that aging scientists are actively fighting against change? I do not doubt that the three measured factors are causally related to disruption, and co-related to one another, but the interpretation seems to miss within-field and within-person forces that do not diminish their importance, but change their interpretation and neutralize the associated policy implication (write less, read more—P. 7—makes less sense unless the writing less prescription is redefined to be not over historical time but over the research career; and out of field). I note that I did find the observation of a conservation of disruptive papers and patents a very interesting finding, mic-dropped at the end, but would have loved for this to have been detailed/demonstrated/interpreted further.

I did not fully understand or was persuaded by the author’s demonstration that “this trend is unlikely driven by change in citation practices or the quality of published work.” The first is supposedly demonstrated by rewiring citations to preserve network structure and age and showed a similar CD actual/rewired ratio (hovering, fluctuating around .75, but dropping for physical science and rising above 1 for electrical and electonic patents). But the change if the change in citation patterns has to do with an increase in relevant past work (which it does) or the change in network structure (e.g., entropy, which it also does) then what other style of citation change was this supposed to exclude? The quality of published work shows that papers in top outlets manifest the same as all other work, which suggests that this is a secular change, which may be about decreasing innovation, but may also be about prior relevance.

What was notably absent here was discussion/measurement of one of the many operationalizations of combinatorial novelty (Uzzi et al. 2013; Sorenson, Rivkin, and Fleming, n.d.) over time, despite efforts to measure the described performance of novel things (e.g., the semantic categories of “creation” and “discovery”). Obviously, with a greater number of things studied, opportunities for combinatorial novelty increase over time and so this is less likely to drop in the same way that disruption does, but we should know this based on its importance to advance and its clear relation to creation, discovery and invention in both science and technological domains.

Still and all, the paper boldly documents a big trend (caveats as stated above) and is clearly written (despite incompleteness in the math of their measure and the procedure for their survey, which could trivially be fixed). A mild, final concern relates to the theme of slowing science, which has become something of a commonplace in science, based on several studies they cite (Bloom et al. 2020; Jones 2009; Chu and Evans 2021; Jones and Weinberg 2011; Wu, Wang, and Evans 2019; Milojević 2015), but also editorialized more broadly (e.g., “Science is Getting Less Bang for Its Buck” by Collison and Nielsen in the Atlantic). What exactly do the authors contribute to this narrative. They show that there is more cited work in the past perceived as relevant, more uneven/unequal attention to that work (“diversity”), and fewer works seen as framing new directions with the policy prescription to write less and read more. This work contributes to the narrative, but it doesn’t change it. To be clear, however, it is an important narrative—maybe the most important for the Science of Science and Innovation—and one worth understanding better. Moreover, there were hints of things that do or could contribute in new ways, like the seeming conservation of disruption mentioned in the last paragraph that could suggest something of a carrying capacity in science.

Detailed/minor comment:
Figure 1. Fails to describe the function in C: no definition of f_it, b_it or how the measure works despite its (uninterpretable) illustration. Also, it describes CD_t=0 as its median value, but this, obviously hinges on the distribution of CD_t values and relates to my earlier main point about a lack of care regarding CD distributions that are neither described nor graphed and illustrated (except for

here, with the almost assuredly false characterization that 0 is median).

Extended Data Fig. 1A is broken: the measurements prior to 1992 to appear to missing…maybe from a rendering error (but is this supposed to be the same as Figure 3A?…I think so, but then why do 3B and ED1B not appear to line-up?)

SI Line 593: “A small number of words APPEARED in disruptive”

SI heading 3 “Classification of disruptive and consolidating words”. The authors mentioned an “All Our Ideas” survey and the three people who filled it out,

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. 2020. “Are Ideas Getting Harder to Find?” The American Economic Review 110 (4): 1104–44.

Chu, Johan S. G., and James A. Evans. 2021. “Slowed Canonical Progress in Large Fields of Science.” Proceedings of the National Academy of Sciences of the United States of America 118 (41). https://doi.org/10.1073/pnas.2021636118.

Jones, Benjamin F. 2009. “The Burden of Knowledge and the ‘Death of the Renaissance Man’: Is Innovation Getting Harder?” The Review of Economic Studies 76 (1): 283–317.

Jones, Benjamin F., and Bruce A. Weinberg. 2011. “Age Dynamics in Scientific Creativity.” Proceedings of the National Academy of Sciences of the United States of America 108 (47): 18910–14.

Milojević, Staša. 2015. “Quantifying the Cognitive Extent of Science.” Journal of Informetrics 9 (4): 962–73.

Sorenson, Olav, Jan W. Rivkin, and Lee Fleming. n.d. “Complexity, Networks and Knowledge Flow.” SSRN Electronic Journal. https://doi.org/10.2139/ssrn.310001.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. “Atypical Combinations and Scientific Impact.” Science 342 (6157): 468–72.

Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. “Large Teams Develop and Small Teams Disrupt Science and Technology.” Nature 566 (7744): 378–82.

Referee #3 (Remarks to the Author):

This is a novel contribution to an important conversation about whether the research enterprise is generating less progress despite the strong growth in production of papers. The authors focus on a metric measuring the extent to which papers do not also cite the references of the papers they reference. The logic is that if a paper is a breakthrough, it shifts the conversation into new directions and papers that build on the breakthrough paper no longer need to pay attention to what the authors of the breakthrough paper were attending to. On the other hand, if a paper adds to an ongoing conversation, then papers that build on it will be motivated by the same work and will therefore cite the same work. The rationale makes sense to me.

Other papers have opened up the possibility that progress is slowing, but as the authors state the basis for those arguments has been field specific. This paper tackles the question across the full research domain. The OECD is concerned about the problem. They are producing an edited book on the topic and whether AI can offer solutions based on summer conference. I take this as evidence that this is a high level problem of international concern.

A more nuanced reading is warranted of Figure 2: Decline of disruptive science and technology. This is the display of the main result, the declining metric of disruptiveness. The paper figure begins in 1945 and the interpretation in the text reports decline since 1945. But is it really surprising that the world today differs from the world in 1947? Not so much. If we look at the situation in papers since 1980, physical sciences has been stable and life sciences pretty much so except for a step change (which I never trust in such data) in about 1997; WoS may have changed its coverage in 1997. Social sciences shows the most dramatic and continuous drop since 1980. The disruptiveness in patents has been stable for about the last quarter century, i.e. since about 2000.

I recommend dropping "We verify the decline in disruptiveness over time through regression models, which show that publication year has a statistically significant and negative relationship with CD5 for papers ($p < 0.001$) and patents($p<0.001$)." The drop visible in the graph does not need verifying. Instead, more subtlety in interpretation is required.

I would like more clarity on some aspects of method.

1 On alternative samples. As written the text implies analysis was run on JSTOR and APS. But that is impossible. The metric concerns citing and referencing. JSTOR and APS do not index references and citations. Was the analysis performed on records indexed in WoS that are also indexed in JSTOR and APS, and PUBMD? In fact, was the Microsoft Academic Graph analysis also done by extracting records from MAG and then looking at the metrics calculated in WoS? That would require less effort so I assume that was how it was done rather than rebuilding the system in MAG. The text should describe exactly what was done, including how many items from each source could not be found in WoS.

2 On the Monte Carlo simulation. The origin of the metric lies in network analysis. As network analysts, the authors find convincing a network analysis technique of comparing the data to a randomly rearranged network. However, not being a network analyst, I don't see the rationale here. It would seem that all that is achieved by comparing it to a random graph is establishing that it is not random. We know that the citation graph is not random. We don't need a Monte Carlo simulation to establish that.

I would like to see that dropped and replaced with a regression that establishes the extent to which the change in the metric over time is explained by change in number of papers published, number of references per paper, and number of authors per paper. The very last item in the supplementary information does part of this. So I am saying move it to the main text and include trends in number of references per paper and authors per paper, both of which could be plausibly related to trends in the metric. This is not the same as extended table 2 which seeks to explain a paper's metric using characteristics of that paper.

On details of presentation. I was somewhat frustrated trying to piece together the material spread across manuscript with references, methods with references, extended data and supplementary information. However, the fault lies not with the authors but with Nature not being able to afford social science papers the room they need to explain their material thoroughly. Anyway, some details need cleaning up.

- Extended figure 1 is not mentioned in the extended text. It feels like it is hanging out there by itself. It should be mentioned in the extended text.
- Please use the non-jargon explanation of lexical diversity – unique/total words because without the explanation (provided in the text but not the extended text) the jargon will be obscure to many readers
- Extended figure 2 – We know papers have increased a lot. Do not plot them here (blue line) because it so compresses the y axis that we cannot effectively judge the claims about the papers of interest, those with CD>0.25.
- Supplementary information, lexical diversity – Was uniqueness defined within a field or across the whole dataset? How do we know that even obsolete words are occasionally used? Is there a reference? Your analysis? Obsolete words are used in titles and abstracts? The second sentence in this section talks about increases in disruption, but the paper searches for decreases in disruption. Change "increases" to "decreases".
- Supplementary table 1 – The table displays verbs. The table note is all about nouns. Replace current note with explanation of the colors of the words and include black in that explanation. Also explain how the numbers were calculated
- On English usage, much use is made of "utilize". Use is preferable.

Author Rebuttals to Initial Comments:

Response to Reviewer #1
(*Original Comments in Italics* / Author Responses in Standard Typeface)

R1 General: *Using large-scale datasets, the authors investigated whether progress in science is slowing in major fields or not. The analyses are mainly based on the newly introduced disruption indicator. The study seems technically and methodologically sound.*

Response: Thank you for your careful engagement with our work and for providing a succinct overview. We appreciate all of your thoughtful suggestions for improving the manuscript. As you will see below, we have tried our best to fully respond to the feedback in your review. We hope that you find the revised manuscript to be a significant improvement over the previous version.

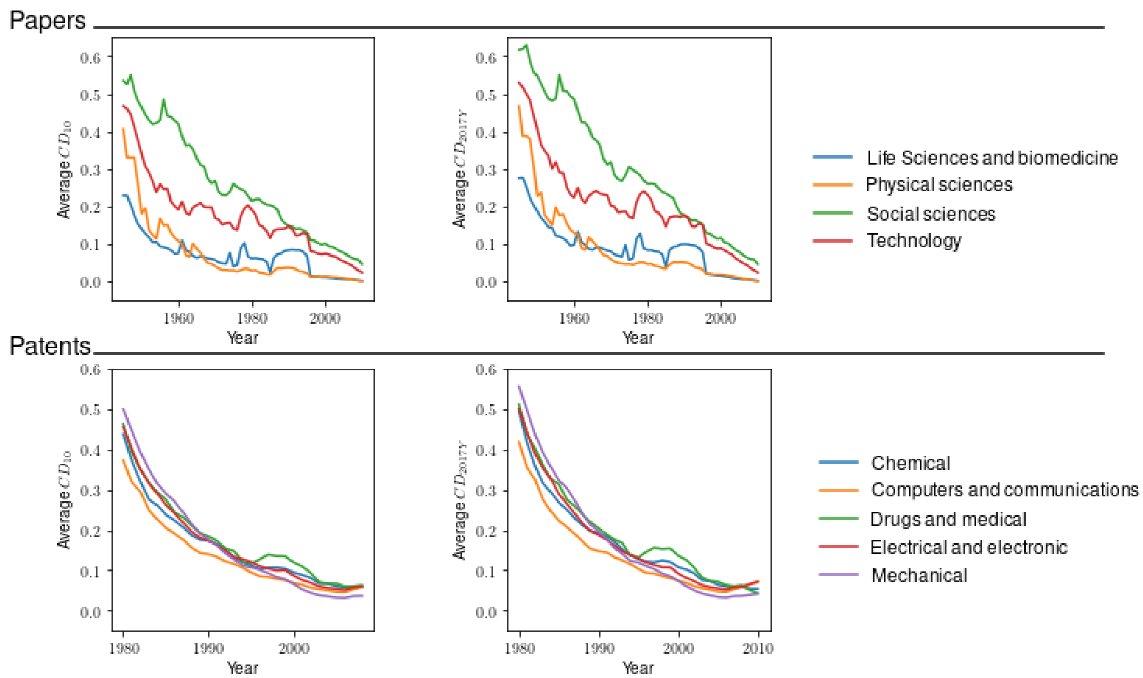
R1C1: *Lines 61-62: It is not clear what the percentages mean. Why didn't the authors present mean CD index values instead?*

Response: Thank you for drawing our attention to this lack of clarity in our presentation. In the updated manuscript, we revised our discussion to present mean values of the CD index. We also include the percentages for readers who may want to compare the relative decrease across fields, along with some additional explanation to clarify the meaning of the percentages (see p. 4 lines 63-68).

R1C2: *Furthermore, I am not sure whether CD index values can be compared across decades. The publication and citation cultures (and database coverages) are so different that these comparisons might be impossible. In bibliometrics, it is conventional to time-normalize indicators – besides field-normalization (Waltman, 2016). I propose to do the same with disruption indicators (see Bornmann & Tekles, 2019).*

Response: Thank you for drawing our attention to this important issue. In light of your feedback, we now appreciate that in the prior version of our manuscript, we had not sufficiently addressed potential concerns about use of the CD index over the long time period spanned by our study. In the revised paper, we take a multi-pronged approach to address this critical issue, including assessments of sensitivity with respect to the measurement window and adjustment for potential confounders using regression, normalized indicators, and computer simulation. As we detail below, the results of these assessments and adjustments give us confidence that our results are unlikely due to changes in publication, citation, authorship, or other potential confounders.

1. First, based on your comment and the insight from Bornman and Tekles (2019) we computed the CD index with longer time-windows and included the plots in the manuscript (p. 32 Extended Data Figure 2), which are also shown below.

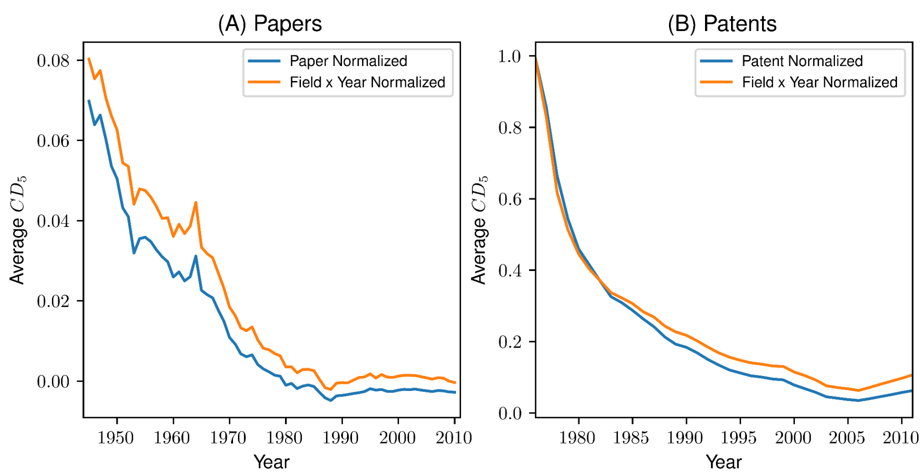


The left hand column shows the results for the CD index measured based on citations made within 10 years of the focal paper's (or patent's) publication (or grant) year. The right hand column shows the results for the CD index measured based on all citations made between the focal paper's (or patent's) publication (or grant) year and 2017. In the plots for CD_{10} (left column), we see the decline persist when we take into account the citation patterns of works that are published within 10 years of the focal work's publication or grant year. In the plots for CD_{2017} (right column), the pattern holds when the forward citations of works published through 2017 are accounted for by the measure. Put differently, even when taking into consideration that some papers and patents may take longer than five years for their forward citations to exhibit a stable level of disruptiveness (Bornmann and Tekles, 2019), we still document a decline in disruption. Moreover, a comparison across the two columns (and with the plots in Figure 2 of the main text) indicates that the decline is somewhat steeper when we use longer forward citation windows, suggesting that the results we present in the main text (which are based on a 5 year window) may represent conservative estimates of the decline. In short, these results suggest the decline in the CD index is unlikely to be driven by our choice of time-window for computing the measure. Accordingly, we have mentioned the robustness of our results to varying time-windows and have cited the key paper (e.g., Bornmann and Teckles 2019) that suggests this possibility (p. 3 lines 57-58).

2. Second, following your suggestion, we devised several normalized versions of the CD index that aim to better facilitate comparisons over time (p. 21 lines 407-433). In developing these metrics, we found the Waltman (2016) article mentioned in your review to be extremely helpful, along with related papers by Bornmann and Marx (2015), Petersen et al. (2019), Waltman and van Eck (2019), and Bornmann (2020) (all of which are now cited in the paper). Among the various components of the CD index, we focused our attention on the count of papers or patents that only cite the focal work's references ("N_k"), as this term would seem most likely to scale with the increases in publishing and patenting and in the average number of citations made by papers and patents to prior work (c.f., Bornmann et al., 2020). Larger values of N_k lead to smaller values of the CD index. Consequently, dramatic increases in N_k over time, particularly relative to other components of the measure, may lead to a downward bias, thereby inhibiting our ability to accurately compare disruptive science and technology in later years with earlier periods.

Our two normalized versions of the CD index aim to address this potential bias by attenuating the effect of increases in N_k. In the first version, which we call **"Paper normalized,"** we subtract from N_k the number of citations made by the focal paper or patent to prior work ("N_b"). The intuition behind this adjustment is that when a focal paper or patent cites more prior work, N_k is likely to be larger because there are more opportunities for future work to cite the focal paper or patent's predecessors. This increase in N_k would result in lower values of the CD index, though not necessarily as a result of the focal paper or patent being less disruptive. In the second version, which we call **"Field × year normalized,"** we subtract from N_k the average number of backward citations made by papers or patents in the focal paper or patent's WoS research area or NBER technology category, respectively, during its year of publication (we label this quantity "N_b^{mean}"). The intuition behind this adjustment is that in fields and time periods in which there is a greater tendency for scientists and inventors to cite prior work, N_k is also likely to be larger, thereby leading to lower values of the CD index, though again not necessarily as a result of the focal paper or patent being less disruptive. In cases where either N_b or N_b^{mean} exceed the value of N_k, we set N_k to 0 (i.e., N_k is never negative in the normalized measures).

In Extended Data Figure 10 (p. 40), also pasted below for convenience, we plot the average values of both normalized versions of the CD index over time, separately for papers (Panel A) and patents (Panel B). Consistent with our findings reported in the main text, we continue to observe a decline in the CD index over time, suggesting that the patterns we observe in disruptive science and technology are unlikely driven changes in citation practices.



Additionally, to complement our two normalization approaches, we also computed an alternative, previously developed variation on the CD index that is less susceptible to changes in citation practices over time. Specifically, as discussed in the manuscript under "Alternative bibliometric measures" (p. 20 lines 379-388), Bornmann and colleagues (2020) proposed a novel disruption indicator, $DI_l^{no\ k}$, that, by design, does not include the N_k term, and that therefore may enable more robust comparisons in the disruptiveness of papers and patents across time periods. As we show in Extended Data Figure 9 (p. 39), we find that the average value of $DI_l^{no\ k}$ declines over time, similar to what we see using the CD index.

3. Third, following suggestions from Reviewer #3 **(R3C5)**, we now conduct a series of regression analyses to examine whether the observed declines in the CD index are attributable to changes in publication, citation, or authorship practices (for details, see p. 21-22 lines 434-457). Specifically, in the updated manuscript, we report the results of regression models (p. 51 Supplementary Table 2)

predicting the CD_5 index of papers (Models 1-4) and patents (Models 5-8), with indicator variables included for each year of our study window. The most conservative estimations (Models 4 and 8 for papers and patents, respectively) include subfield fixed effects and control variables for several field × year level—*Number of new papers/patents*; *Mean number of papers/patents cited*; *Mean number of authors/inventors per paper*—and paper/patent-level—*Number of papers/patents cited*—characteristics, thereby enabling more robust comparisons in patterns of disruptive science and technology over the long time period spanned by our study. Across models, we find that the coefficients on the year indicators are statistically significant and negative, and growing in magnitude over time (see p. 41 Extended Data Figure 11 for a visualization of the model predictions), consistent with the patterns we reported based on unadjusted values on the CD_5 index in the main text.

4. Finally, we considered whether our results may be an artifact of changing patterns in publishing and citation practices by using a simulation approach. While this approach was included in the prior manuscript, we agree with the review team that our presentation was difficult to follow. In the revised manuscript, we have completely reworked the motivation, approach, and interpretation for the simulation analyses, and tried to more clearly articulate how the findings add confidence to our results. For a more detailed discussion, please see p. 22-23 lines 458-502.

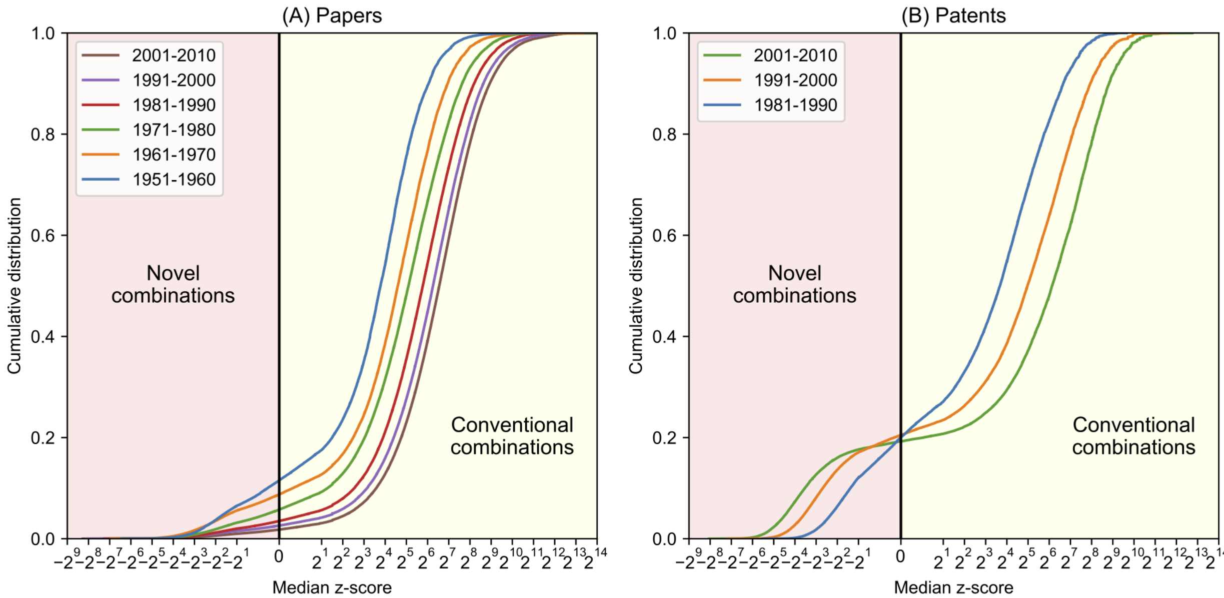
Once again, we would like to thank you for bringing this important issue to our attention, and for your concrete and helpful suggestions for improvement. We hope that the additional steps we have undertaken address your concerns.

R1C3: *Lines 69-71: This analysis could be connected to the incommensurability hypothesis of Thomas Kuhn (Wray, 2011).*

Response: Thank you for bringing this literature to our attention. We added a citation to this stream of literature and also highlight the connection to Kuhn's and Wray's work (p. 4 lines 76-78).

R1C4: *Further indicators that could be used for robustness checks in the study are novelty indicators (Wang, Veugelers, & Stephan, 2017). The development of these indicators has a long tradition in scientometrics, and the indicators are conceptually close to disruption.*

Response: Thank you for this suggestion. Reviewer #2 also made this point (please see **R2C8**) and specifically suggested using the novelty measure developed in Uzzi et al. (2013). Incidentally, Wang, Veugelers, and Stephan (2017) state that their measure is conceptually consistent with the measure of Uzzi et al. (2013). Specifically, they write "...we follow Uzzi et al. (2013) in viewing journals as bodies of knowledge and construct our novelty measure based on new combinations of journals in the references" (p. 1424). In light of these considerations, we computed Uzzi et al.'s measure of atypical combinations for both papers and patents. For papers, our data (Web of Science) are identical to Uzzi et al.'s and we were therefore able to follow their approach exactly. For patents (which were not studied in the Uzzi et al. paper), we adapted the Uzzi et al. approach to examine pairings among the primary USPC classes of the cited patents (rather than journals). We present the results of this analysis following the approach of Uzzi et al., who plotted the cumulative distribution function of their novelty measure, with separate lines for subsets of interest. In our case, we plot separate lines for papers and patents published in each decade of our study window. The results are shown below and also included in the revised manuscript (p. 34 Extended Data Figure 4).



Overall, we find a rightward shift in the cumulative distributions over time, suggesting that for both papers and patents, combinations are increasingly more conventional than would be expected by chance, consistent with what we would anticipate based on our results using the CD index. Interestingly, for patents, there is also a smaller shift in the opposite direction on the left side of the distribution, suggesting that novel patents in recent decades are somewhat *more* novel than novel patents in earlier decades. Still, the bulk of the distribution is moving to the right, indicating greater conventionality.

In addition to showing changes in combinatorial novelty using Uzzi et al.'s (2013) measure, we also added a text-based analysis of the novelty of word pairs appearing in the titles of papers and patents over time (see Figure 3 for the results; our methodological approach is described in Section 1 of the Supplementary Information [p. 47-48 lines 678-683]). Consistent with our findings using Uzzi et al.'s (2013) measure of combinatorial novelty, we observe a persistent decline in new word pairs over time.

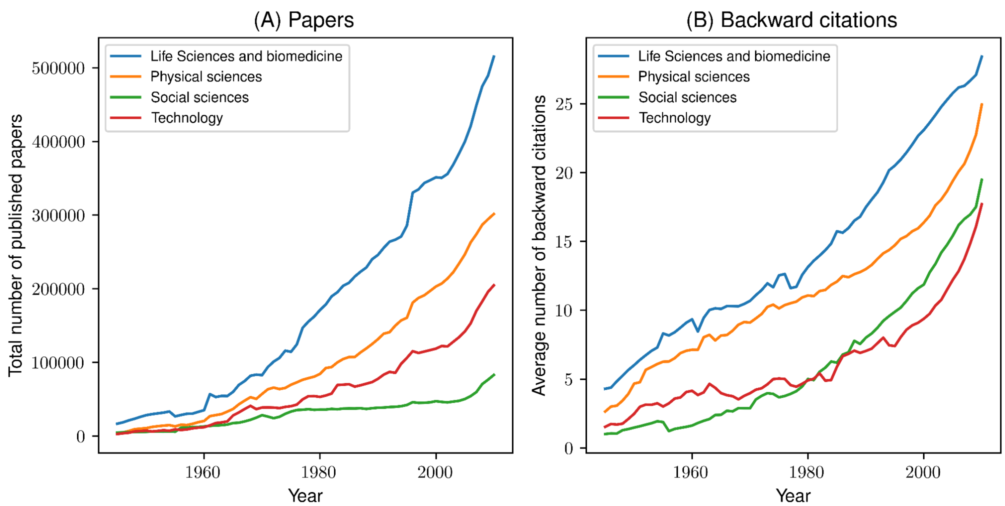
R1C5: *Lines 84-85: In my opinion, "produce" is not a verb with a (close) link to disruption. I cannot follow the rational of using this word as a link to disruption. The same is true for the word "use" and its link to consolidating.*

Response: Thank you for drawing our attention to these potential points of confusion. In light of this comment, along with similar ones from other members of the review team, we can now see that our verb classification and coloring exercise had some limitations and ultimately added more confusion than clarity, and therefore we dropped it. Instead, in the revised manuscript, when we discuss changes in the use of verbs in paper and patent titles, we simply provide suggestive interpretations of the most conspicuous patterns (p. 47-48 lines 711-719).

R1C6: *Figure 2: I see sharp increases of CD valuesin 1960 for social sciences and around 1980 for technology. Please explain these increases.*

Response: Thank you for your careful reading of our figures. We agree that Figure 2, although generally showing a decline in the average value of CD_5 over time, has interesting periodic fluctuations. Based on your feedback, we dug more deeply into the increases in 1960 for social science and 1980 for technology. The aim of this further exploration was to better understand what may be driving these intriguing patterns. First, we wanted to verify the observed fluctuations were not due to technical errors (e.g., anomalies in the underlying data) that distorted the average value of CD_5 index. Second, we tried to better understand the timing and scope of the increases to identify some of their potential substantive drivers.

1. To begin, we wanted to verify that the temporary increases were not due to data issues. In particular, the journals that were indexed by Web of Science may have shifted during the period around the fluctuations in question, leading to changes in the structure of the citation data, which may in turn account for various local peaks in the CD_5 index when plotted over time. If this were the case, we would expect to see proportional shifts in the number of papers that are covered by Web of Science and/or changes in the number of backward citations made by papers during the periods of fluctuation. Based on these considerations, we plotted the number of papers and the average number of citations made by papers in Web of Science (see below).



In the figure above, we find little evidence of proportional fluctuations in the number of papers or citations in social science around 1960 or technology around 1980. These patterns are generally consistent with prior work that has verified the reliability of Web of Science data in the postwar era (Adriaanse and Rensleigh, 2013, de Winter et al., 2013, Singh et al., 2021). In short, we do not find substantial evidence to suggest that the temporary increases in the CD_5 index are caused by data errors.

2. Given that the fluctuations do not seem to result from changes in the data structure, we performed a deeper dive to see whether there may be a substantive explanation. We pinpointed that the year-on-year increase occurred between 1954 to 1956 for social science papers and between 1976 to 1979 for technology. Thereafter, we sought to determine whether changing patterns of discovery (e.g., an emerging school of thought or a salient new technology) may be coincident with the fluctuations. To do so, we disaggregated the data to the WoS subject area level (WoS "Subject Areas" are subsets of larger WoS "Research Areas") to identify those in which the average CD_5 was increasing the most in the years in question for social science and technology. Table A shows the five subject areas that experienced the greatest increase over the prior year for 1955 (first three columns) and 1956 (last three columns) within the social sciences. Table B shows the five subject areas that experienced the greatest increase over the prior year for 1977 (first three columns), 1978 (middle three columns), and 1979 (last three columns) within technology.[1] We note that with the exception of

[1] To provide some specific examples of disruptive papers represented in these tables, "The meaning and assessment of intelligence" (Burt, 1956) was published in the subject of "Biomedical Social Sciences," which is the subject that underwent the greatest increase in the CD_5 index within the social science research area in 1956. The paper represented a departure from existing studies on intelligence by exploring the concept of human intelligence from a eugenics perspective, using ideas from philosophy, biology, physiology, individual psychology, and statistics. Accordingly, its CD_5 value is 0.33. Additionally, "A computer oriented approach to the comprehensive organization of information in hand-compiled medical records" (Certttti et al., 1979) was

"Biomedical Social Sciences," no subject area appears more than once within each table. The wide range of subject areas that are contributing to the observed increases in the social sciences and technology suggest that the temporary divergences from the overall decline in disruptiveness are likely the result of scattered spurts of disruptive works across various fields during these time periods.

Table A: Subject areas in Social Science with highest increases in the CD_5 index from 1955-1956

Year	Subject area	CD_5 increase	Year	Subject area	CD_5 increase
1955	Criminology & Penology	0.66	1956	Biomedical Social Sciences	0.58
1955	Biomedical Social Sciences	0.32	1956	Archaeology	0.45
1955	Social Issues	0.23	1956	Operations Research & Management Studies	0.32
1955	Archaeology	0.19	1956	Social Work	0.19
1955	Anthropology	0.12	1956	Geography	0.17

Table B: Subject areas in Technology with highest increases in the CD_5 index from 1977-1979

Year	Subject area	CD_5 increase	Year	Subject area	CD_5 increase	Year	Subject area	CD_5 increase
1977	Communication	0.66	1978	General & Internal Medicine	1.02	1979	Health Care Sciences & Services	0.50
1977	Linguistics	0.50	1978	Philosophy	0.93	1979	Mathematical Methods in Social Sciences	0.45
1977	Public Administration	0.50	1978	Transplantation	0.50	1979	Meteorology & Atmospheric Sciences	0.30
1977	Marine & Freshwater Biology	0.39	1978	Mining & Mineral Processing	0.36	1979	Physical Geography	0.27
1977	Literature	0.25	1978	Agriculture	0.33	1979	Behavioral Sciences	0.24

We would like to thank you for this comment, which pushed us to think more carefully about the evidence we present in the paper. In the process of addressing this comment, we failed to find patterns that lead us to question the validity of our data or results. Subsequent qualitative research which focuses on local trends within social science or technology may be needed to assess potential drivers of increasing disruption during these periods. Nevertheless, we hope that our response here addresses your concern.

R1C7: *Lines 216-217: This paper has been published (Leydesdorff, Tekles, & Bornmann, 2021).*

Response: Thank you for bringing this to our attention. We have now updated the citation and its reference to reflect the latest status of the manuscript (p. 6 line 128).

R1C8: *Lines 304-305: What is the reason for the random selection?*

Response: Thank you for drawing our attention to this lack of clarity in our presentation. As we now discuss in the revised manuscript (p. 20 lines 366-375), the Microsoft Academic Graph (MAG) data are exceptionally large relative to the other data sources, which imposes a significant computational burden. In light of this and because our analysis of the MAG data was intended to corroborate our findings using WoS and PatentsView, we decided to compute the CD index on a random sample of the data rather than the entire Microsoft Academic Graph.

---

published in the subject of "Health Care Sciences & Services", which is the subject that experienced the largest increase in the CD_5 index within the research area of technology in 1979. The paper explored the increasing use of electronic medical record keeping, which was a new technology that departed from traditional record-keeping practices. Appropriately, the paper had a CD_5 value of 0.25.

R1C9: *Lines 420-422: A paper with more recent results exist (Bornmann, Haunschild, & Mutz, 2021).*

Response: Thank you for directing us to this relevant and recent paper. We have now included a citation to this work in the revised version of the paper (p. 6 line 131).

R1C10: *Page 26: This is wrong: "For papers, lines correspond to WoS research areas; for patents, lines correspond to NBER technology categories".*

Response: Thank you for this comment. We have now edited the caption to correctly reflect the figure (p. 35).

R1C11: *Page 28: Figure B is especially interesting. Until 1960, mean DI values are (significantly) higher than 0. Since then, they are around zero. Furthermore, papers with DI = 1 are especially visible before 1960. I wonder how these results can be explained. Is there a change in the structure of citation data?*

Response: Thank you for your careful attention to our results. We agree that the concentration of highly disruptive Nobel Prize winning papers in the earlier decades of the 20th century is interesting and worthy of further investigation.

To begin, we note that Extended Data Figure 7B (p. 37) extends much further back in time (to 1900) than our main analyses (which begin in 1945). Although Web of Science continues to be regarded as one of the most exhaustive (Singh et al., 2021), reliable (Adriaanse and Rensleigh, 2013), and expertly-curated (de Winter et al., 2014) sources of scientific publication data (Fortunato et al., 2018), its coverage and accuracy is likely to be less robust in earlier eras (a point we now note in the revised caption for Extended Data Figure 7B), and as your comment intimates, the relative sparsity of records from earlier years may potentially be a reason why we see the visual discrepancies in the number of highly disruptive Nobel papers. These considerations of data quality, along with important science policy changes during the postwar era (e.g., Kevles 1977; Appel 2000) **are why we start our main analysis in 1945,** which is consistent with many other Science of Science studies that use data from Web of Science (e.g., Wutchy et al., 2007; Jones et al., 2008; Wu et al., 2019).

With that caveat in mind, because the Nobel Prize data extend back to 1900 (Li et al., 2019), the first year in which the Nobel Prize was awarded, we thought it may be of interest to some readers to present the full time series. We note that even looking to the period of 1945-2010 (the period covered by our main analysis) we continue to detect a persistent decline in disruptiveness among Nobel papers. This suggests that the decline in the disruptiveness of works associated with Nobel prizes is unlikely to be driven solely by the robustness of Web of Sciences's coverage in the prewar era.

From a more substantive point of view, we are not the first to suggest a decline in the disruptiveness of Nobel prize-winning papers. Others have made similar observations at a smaller scale (e.g., Collison and Nielsen 2018; Cowen and Southwood 2019; Cowen 2020). In particular, prior studies have noted that in comparison to earlier Nobel prize-winning papers, recent winners in the fields of physics, medicine, and chemistry (the fields plotted in Extended Data Figure 7B) feature fewer "fundamental breakthroughs" (Cowen and Southwood 2019, p31). Therefore, although we are among the first to quantitatively document the decline in disruptiveness of Nobel prize papers at scale using CD_5, the trend is consistent with existing reports.

Overall, we believe the trend we document among Nobel associated papers is likely related to reports of a decrease in the "boldness" of Nobel works (Collison and Nielsen 2018; Cowen and Southwood 2019; Cowen 2020). More importantly, the decline in disruptiveness among Nobel papers indicates that the trend is independent of the quality of papers—which we further verify in the accompanying figure (Extended Data Figure 7A)—and as we argue, likely represents a sweeping trend in science. We thank you for engaging closely with this specific evidence we present and encouraging us to think more deeply about it.

R1C12: *Page 30: Several CD index variants are discussed in both papers. Please explain which have been used in the study. Give reasons why you preferred certain variants.*

Response: Thank you for pointing this out. We have now identified which variants were used and added a discussion of why we chose these particular measures in the Methods section (p. 20 lines 376-388).

R1C13: *Section 3: Here, it is not clear what the authors did. Why did they involve two researchers who independently searched for words? Furthermore, the survey is not clear. How many people have been asked, and answered? Who has been asked? Researchers? What kind of sample has been surveyed?*

Response: Thank you for this comment. As we explained above in response to your point **R1C5**, in light of this comment, along with similar ones from other members of the review team, we can now see that our verb classification and coloring exercise had a number of significant limitations and ultimately added more confusion than clarity, and therefore we dropped it. Instead, in the revised manuscript, when we present changes in the use of verbs in paper and patent titles, we simply provide suggestive guidance on how to interpret some of the most conspicuous patterns (p. 47-48 lines 711-719).

Response to Reviewer #2
(*Original Comments in Italics* / Author Responses in Standard Typeface)

R2 General: *In "the decline of disruptive scienceand technology" the authors use multiple measures, including an interesting one that they recently invented (and has received widespread attention),to trace a declining trend in science and technology over the latter half of the 20th Century. They analyze this in the context of citation and word-use patterns across multiple datasets, attempt to control for confounding explanations, and try to identify explanations, painting a coherent portrait of science and technology institutions that have slowed as they have grown, aged and narrowed.*

There are several things to like about the manuscript, including the careful, redundant measurement across fields (and even domains, e.g., JSTOR vs. the Web of Science us. USPTO). Nevertheless, I had some concerns about the manuscript and its claims with respect to measurement, effect identification, interpretation and framing.

Response: Thank you for your summary of our work and for pointing out some of its key strengths and weaknesses. We appreciate all of your thoughtful suggestions for improving the manuscript. As you will see below, we have tried our best to fully respond to the feedback in your review. We hope that you find the revised manuscript to be a significant improvement from the previous version.

R2C1: *On measurement, the Consolidating/Disrupting (CD) measure, which forms the core of the author's empirical argument is based on a quantity that necessarily decreases with the amount of science available to be cited (the "b" of prior works cited--a variable stated in Figure 1, but not defined across the paper). In other words, while their interpretation of the measure as an indication that a word is viewed as representing a new direction in science or technology is valid at any given point of time (e.g., for two focal pieces at any given time), its comparison across time is suspect because the amount of science at risk of being cited, and relevant to current science is increasing over time, which can only trivially be interpreted as a reduction in innovation. In other words, with more work being cited earlier because there is more relevant work to be cited (and an aging scientific workforce that is citing older work), the disruption measure feels less like an indicator of new views in science emerging over time, than a shift in the measurement of scientific views. In other words, there are more (citational) stories to tell about how recent work is tied in with the past over time. The authors take pains to reduce the historical variation in the measure by only looking at the citations 5-years out, but they cannot as easily account for the changing relevance of the past. Incidentally, the same historical pattern is demonstrated in Extended Data Fig. 3 of (Wu, Wang, and Evans 2019) in each panel c through m. The authors state that "these declines demonstrate that relative to earlier eras, recent papers and patents do less to push science and technology in new directions". This feels more directly demonstrative that there are more possible stories of progress—where the present builds on the past—to tell about any particular recent paper.*

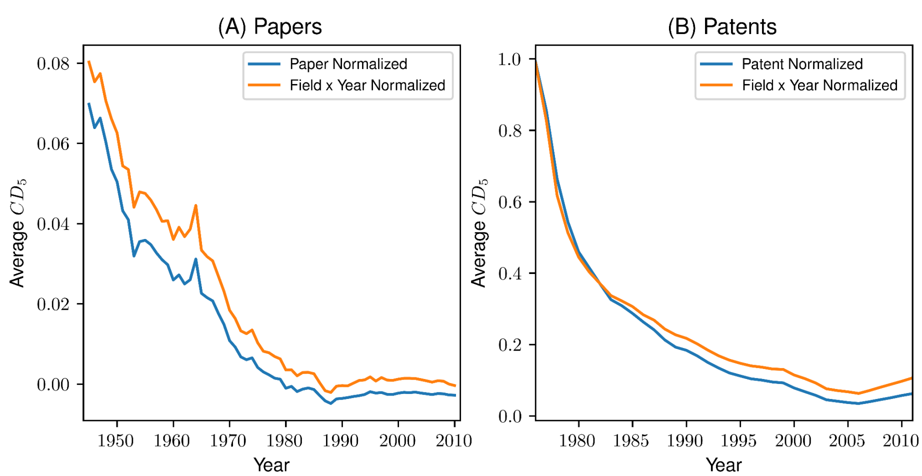
Response: Thank you for your careful attention to this important issue. In light of your feedback (and similar comments from Reviewers #1 **[R1C2]** and #3 **[R3C5]**), we can now appreciate that in the prior version of the manuscript, we had not done enough to address potential concerns regarding our comparisons of the CD index over the long time period spanned by our study, particularly with respect to the dramatic growth in the amount of science and technology available for citation. In the revised paper, we take a multi-pronged approach to address this critical issue, including adjustment for potential confounding using normalized indicators (suggested by Reviewer #1 **[R1C2]**), regression (suggested by Reviewer #3 **[R3C5]**), and computer simulation. As we detail below, the results of these assessments give us confidence that our findings are unlikely the result of changes in publication or citation practices or other potential confounders.

1. First, following suggestions from Reviewer #1 **(R1C2)**, we developed several normalized versions of the CD index that aim to better facilitate comparisons over time (p. 21 lines 407-433). Among the various components of the CD index, we focused our attention on the count of papers or patents that only cite the focal work's references ("N_k"), as this term would seem most likely to scale with the increases in publishing and patenting and in the average number of citations made by papers and patents to prior work (a point that has been noted in other research on the CD index, c.f., Bornmann et al., 2020). Larger values of N_k lead to smaller values of the CD index. Consequently, dramatic increases in N_k over time, particularly relative to other components of the measure, may lead to a

downward bias, thereby inhibiting our ability to accurately compare disruptive science and technology in later years with earlier periods.

Our two normalized versions of the CD index aim to address this potential bias by attenuating the effect of increases in N_k. In the first version, which we call **"Paper normalized,"** we subtract from N_k the number of citations made by the focal paper or patent to prior work ("N_b"). The intuition behind this adjustment is that when a focal paper or patent cites more prior work, N_k is likely to be larger because there are more opportunities for future work to cite the focal paper or patent's predecessors. This increase in N_k would result in lower values of the CD index, though not necessarily as a result of the focal paper or patent being less disruptive. In the second version, which we call **"Field × year normalized,"** we subtract N_k by the average number of backward citations made by papers or patents in the focal paper or patent's WoS research area or NBER technology category, respectively, during its year of publication (we label this quantity "N_b^{mean}"). The intuition behind this adjustment is that in fields and time periods in which there is a greater tendency for scientists and inventors to cite prior work (e.g., out of the need to tell more stories of how recent work is tied in with the past), N_k is also likely to be larger, thereby leading to lower values of the CD index, though again not necessarily as a result of the focal paper or patent being less disruptive. In cases where either N_b or N_b^{mean} exceed the value of N_k, we set N_k to 0 (i.e., N_k is never negative in the normalized measures).

In Extended Data Figure 10 (p. 40), also pasted below for convenience, we plot the average values of both normalized versions of the CD index over time, separately for papers (Panel A) and patents (Panel B). Consistent with our findings reported in the main text, we continue to observe a decline in the CD index, suggesting that the patterns we observe in disruptive science and technology are unlikely driven changes in citation practices.



Additionally, to complement our two normalization approaches, we also computed an alternative, previously developed variation on the CD index that is less susceptible to changes in citation practices. Specifically, as discussed in the manuscript under "Alternative bibliometric measures" (p. 20 lines 379-388), Bornmann and colleagues (2020) proposed a novel disruption indicator, $DI_1^{no\ k}$, that, by design, does not include the N_k term, and that therefore may enable more robust comparisons in the disruptiveness of papers and patents across time periods. As we show in Extended Data Figure 9 (p. 39), we find that the average value of $DI_1^{no\ k}$ declines over time, similar to what we see using the original CD index.

2. Second, following suggestions from Reviewer #3 **[R3C5]**, we now conduct a series of regression analyses to examine whether the observed declines in the CD index are attributable to changes in publication, citation, or authorship practices (for details, see p. 21-22 lines 434-457). Specifically, in the updated manuscript, we now report the results of regression models (p. 51 Supplementary Table 2) predicting the CD_5 index of papers (Models 1-4) and patents (Models 5-8), with indicator variables included for each year of our study window. The most conservative estimations (Models 4 and 8 for papers and patents, respectively) include subfield fixed effects and control variables for several field × year level—*Number of new papers/patents*; *Mean number of papers/patents cited*; *Mean number of authors/inventors per paper*—and paper/patent-level—*Number of papers/patents cited*—characteristics, thereby enabling more robust comparisons in patterns of disruptive science and technology over the long time period spanned by our study. Across models, we find that the coefficients on the year indicators are statistically significant and negative, and growing in magnitude over time (see p. 41 Extended Data Figure 11 for a visualization of the model predictions), consistent with the patterns we reported based on unadjusted values on the CD_5 index in the main text.

3. Finally, we considered whether our results may be an artifact of changing patterns in publishing and citation practices by using a simulation approach. While this approach was included in the prior manuscript, the review team highlighted that our presentation was difficult to follow, a point with which we fully agree. In the revised manuscript, we have completely reworked the motivation, approach, and interpretation for the simulation analyses, and tried to more clearly articulate how the findings help to add confidence to our results. For a more detailed discussion, please see p. 22-23 lines 458-502.

In addition, we appreciate you drawing our attention to the fact that we had neglected to clearly define the quantity "b" in the prior version of the manuscript. As we describe in more detail in response to your next comment **(R2C2)**, in the revised manuscript, we have updated Figure 1 to make sure "b" (along with the other components of the measure) are clearly defined.

R2C2: *A further measurement issue was the lack of a clear definition of their CD index (I had to go back to their prior 2016 paper to make sense of the variables used in Figure 1 and the calculation it describes), and the failure to describe and illustrate the distribution of this measure at any one point in time (other than tracing its mean in Figure 2), the distribution of varied measures to show their correspondence (e.g., CD at other times that 5), and their stating (erroneously) that the median value is 0 in Figure 1.*

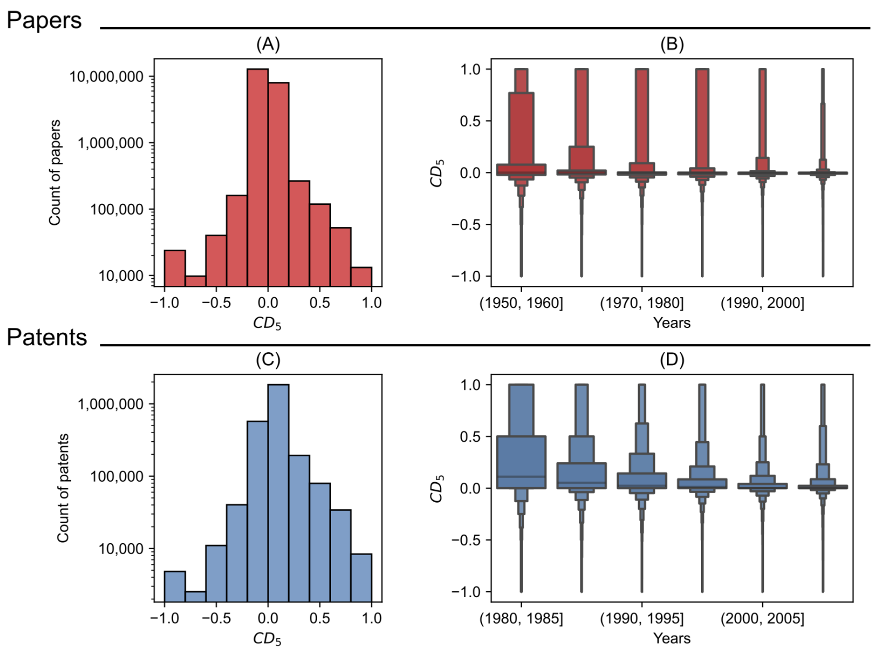
Response: Thank you for drawing our attention to the lack of clarity in our manuscript on these critical measurement issues. Your comment raises several key points, each of which we have addressed in the updated paper.

First, we revised Figure 1 to more clearly define the CD index, including adding explicit explanations of "b_{it}" (i.e., an indicator for whether paper or patent *i* cites the references of the focal paper or patent, mentioned in your previous comment) along with all other variables in the measure.

Second, we have updated the caption to more clearly describe the distributions of the measure shown in Figure 1. Figure 1 Panel B shows the distribution of the CD index for our sample, where the strip on the left represents the distribution of papers in Web of Science and the strip on the right represents the distribution of patents from the USPTO. Each dot represents a single paper or patent. The vertical (up/down) dimension on each strip corresponds to values of the CD index (with axis values shown in orange on the left). The horizontal dimension (left/right) on each strip is only included so that the points do not all overlap (i.e., so we can "jitter" the points). Darker areas on each strip plot indicate denser regions of the distribution.

Third, we updated Figure 1 to describe the illustrative calculation for the citation network where CD_t=0 as the "midpoint" rather than the median; we appreciate you drawing our attention to that inadvertent error.

Fourth, based on your comment, we also dug more deeply into the distribution of the CD_5 index at different time periods, many features of which, as you note, are lost in Figure 2. In particular, we added a new figure (p. 31 Extended Data Figure 1), also shown below for your convenience, which plots histograms of CD_5 for papers and patents along with letter-value plots to illustrate the distribution over time, within 10 (papers) and 5 (patent) year windows.



In the table below, we present additional quantitative summaries of the CD_5 distributions, separately for papers and patents, at the turn of each decade. Similar to what we report in the paper, we see evidence of a decline in both the mean and median CD index over time. We also see declining variance, as evidenced by the smaller standard deviations across decades.

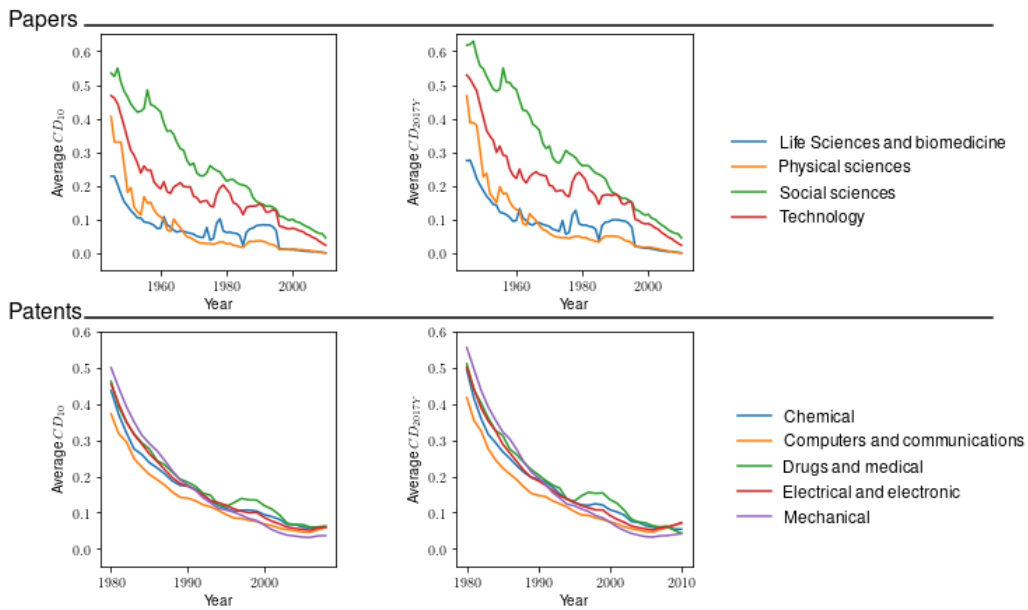
Patents

Grant year	Mean	Median	SD	Min	Max
1980	0.3882	0.2400	0.4586	-1	1
1990	0.1410	0.0370	0.3055	-1	1
2000	0.0638	0.0034	0.2108	-1	1
2010	0.0505	0.0000	0.2024	-1	1

Papers

Publication year	Mean	Median	SD	Min	Max
1950	0.1750	0.0000	0.3964	-1	1
1960	0.1114	0.0000	0.3363	-1	1
1970	0.0697	0.0000	0.2835	-1	1
1980	0.0646	-0.0005	0.2753	-1	1
1990	0.0640	-0.0009	0.2688	-1	1
2000	0.0163	-0.0021	0.1606	-1	1
2010	0.0070	-0.0016	0.1149	-1	1

Finally, following your suggestion (along with a related comment by Reviewer #1, see **R1C2**), we also dug more deeply into the relationship between the CD index measured at t=5 (the forward window we use in our main analyses) and at other forward windows. Specifically, in the updated manuscript, we demonstrate a similar decline in the disruptiveness of papers and patents over time using the CD index measured at t=10 (i.e., 10 years post-publication) and as of the year 2017 (i.e., using all the available citation data) (p. 32 Extended Data Figure 2). We include the figure below for your convenience.



In addition, for further reference, in the table below, we also provide descriptive statistics and correlations for the CD index computed using different forward windows. Overall, we find a high degree of correspondence between the different measures. Indeed, the decline is somewhat steeper when we longer forward windows, suggesting that our main results using a five year window may represent a conservative estimate. These results give us confidence that our findings are not sensitive to our particular choice of forward window.

Distributions of CD index values for different forward windows

Papers

Measure	Mean	Median	SD	Min	Max
CD index measured 5 years post publication	0.0396	-0.0014	0.2209	-1	1
CD index measured 10 years post publication	0.0513	-0.0011	0.2337	-1	1
CD index measured as of 2017	0.0645	-0.0005	0.2453	-1	1

Patents

Measure	Mean	Median	SD	Min	Max
CD index measured 5 years post publication	0.1228	0.0025	0.3030	-1	1
CD index measured 10 years post publication	0.1623	0.0275	0.3266	-1	1
CD index measured as of 2017	0.1713	0.0301	0.3291	-1	1

Correlations among CD index values for different forward windows

Papers

	CD index measured 5 years post publication	CD index measured 10 years post publication	CD index measured as of 2017
CD index measured 5 years post publication	1.000		
CD index measured 10 years post publication	0.9766, p<0.001	1.000	
CD index measured as of 2017	0.9525, p<0.001	0.9801, p<0.001	1.000

Patents

	CD index measured 5 years post publication	CD index measured 5 years post publication	CD index measured as of 2017
CD index measured 5 years post publication	1.000		
CD index measured 10 years post publication	0.9312, p<0.001	1.000	
CD index measured as of 2017	0.8750, p<0.001	0.9428, p<0.001	1.000

R2C3: *To be clear, I liked the word diversity measure (types/tokens) and its consistent pattern over time, which provides a correspondence with the CD results, but the effects become much more modest, especially after 1970 for papers, and 1990 for patents (Figure 3), which are not described within the paper and have an interesting, but undescribed or theorized punctuated form (with a "kink" at 1970).*

Response: Thank you for engaging closely with our empirical results. We agree that there are interesting patterns in the word diversity plots (Figure 3 A and D) that warrant a more nuanced discussion. Thus, when presenting these figures in the revised version of the manuscript, we mention the attenuation in the decline following 1970 for papers and 1990 for patents as you have pointed out (p. 4 lines 81-82).

We do note that the "kink" is most prominent for papers, particularly in the field of Technology. Based on a close investigation of the Web of Science archive documentation along with prior work that has used the data (Adriaanse and Rensleigh, 2013; Clarivate, 2022; de Winter et al., 2013; Fortunato et al., 2018; Singh et al., 2021), we do not believe that a data error driven by coverage changes is responsible for the "kink." Furthermore, when we conducted a detailed analysis of each subject area (i.e., subfield) within Technology, we were not able to identify a compelling substantive change in a particular subfield that would explain the shift. Nevertheless, we agree with you that this is a curious pattern—one that may be usefully investigated in future research. More importantly, although there may be some minor fluctuations, we believe our text-based analyses generally support our main claim that disruption is declining across science and technology.

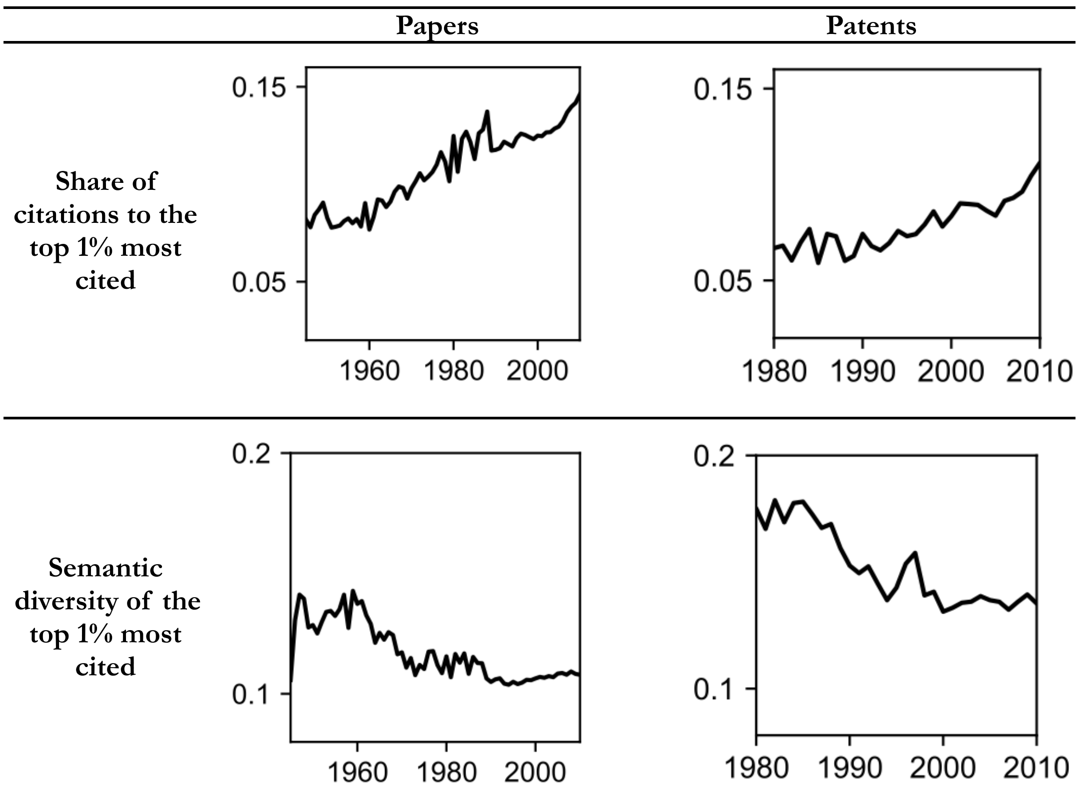
R2C4: *The second analysis of verb classification andscoring was less persuasive to this reviewer. In fact, it was not even totally obvious whether this just one or two separate analyses (described in the section half of SI Section 2 and Section 3). The selection of words seemed idiosyncratic and complex in order to "eliminiate idiosyncratic terms" (SI:506). The procedure involved including verbs (other parts of speech not presented because "verbs...generally yielded more substantively interesting patterns") that significantly correlated with utilization over time, occurred in more than 1000 papers, and excluded "topical terms" that traced the changing "substance" of science rather than its "approach". Then they performed a survey with Science of Science researchers they say is described in Methods (didn't find there, but I did in the next section of the SI...but wasn't sure if that was the same or a different analysis). The survey they described in the next section was performed on the All Our Ideas platform, which fields a "wiki-survey" that performans an pairwise ordinal comparison task (is verb A "more indicative of efforts to create knowledge/technology" than B), sequentially repeated with light (Bayesian) active learning and designed for the crowd-introduction of new topics for comparison. The concern with its use here is that the population of those surveyed was "three researchers who are familiar with the literature on technology strategy and Science of Science." I assume that these are not the three authors on this paper, also presumably familiar with the literature on technology strategy and the Science of Science. This wouldn't necessarily be a problem because the authors chose the 2 sets of categories (about "creation", "discovery" and* *"perception" for disruption, and "assessment", "improvement", and "application" for consolidation), but these are not "validated" by the three researchers, but rather the (tiny) crowd is simply used to source words within each category that are most related to "efforts to create knowledge/technology". There is no indication of how many repetitions the survey was performed; how many unique pairs were floated, or the sequence with which elements were added by the researchers (presumably none were added by the participants, despite this being the primary justification for the active learning approach in All Our Ideas...the survey was designed to converge slowly so that new things could be added and have a chance of ascendancy in the overall score even if added late; there are faster-converging methods when no addition is involved, but we don't know if fast convergence was required because the number of surveyed pairs is not mentioned). Then the top 25 words with the highest score in each disruptive category "were all deemed more likely to appear in disruptive titles" and those highest in each consolidating category "were all deemed more likely to appear in consolidating titles". I'm not sure how this was assessed; maybe from Table 1 (but this focused on the increasing and decreasing terms?) or what statistics were used, or whether or how significantly the associations were. Incidentally, it would have been more persuasive to this reviewer if the survey was used to validate the author's priors about words and word categories and not simply help them characterize their prior. Moreover, the words themselves did not appear semantically distinct or self-describing disruptive innovation vs. consolidation. Take the words in Supp Table 1...its not obvious that "pour", "label", and "note" are more or less disruptive than "sing" "surface" and "disseminate"? "Produce" is the main example from the paper as a disruptive word, but it feels polysemous—are we producing something new for the very first time (but then would we be using the word produce rather than invent or discover?) or are we producing something long-discovered at scale, in which case it becomes conceptually more similar to "automate". In short, (for me) the second semantic analysis created more heat than light.*

Response: Thank you for your careful attention to our methods. In light of your comments, along with similar ones from other members of the review team, we can see that our verb classification and coloring exercise had a number of significant limitations and ultimately added more confusion than clarity, and therefore we dropped it. Instead, in the revised manuscript, when we present changes in the use of verbs in paper and patent titles, we simply provide suggestive interpretations of the most conspicuous patterns (p. 47-48 lines 711-719).

R2C5: *The authors then identify historical decreases in the diversity (normalized entropy) of work cited, increases in the number of self-citations, and increases in the age of cited work. They regress disruption on diversity, self-citation and aging work and find that more diverse, new-to-the-authors, and recent work anticipate disruption (at 5 years) despite changes moving in the opposit direction. I feel that the diversity/disruption findings is an important contribution and demonstration, and it would have been nice for this path-dependant change in the entropy to have been better described (is everyone citing the same things, and are those things semantically like one another?) This was the big take-away for me from the paper.*

Response: Thank you for your kind words on our analysis and suggestions for further development. We agree that the declining diversity of work cited is striking, particularly given the dramatic increase in the volume of published science and technology available for citation. Your specific ideas for unpacking the entropy finding are fascinating—asking whether everyone is citing the same things, and whether those things are more semantically similar—and we implemented both of them in the revised manuscript. To do so, we focused our attention on the top 1% most highly cited papers within each field and year. Here, our rationale was that if scientists and inventors are increasingly citing the same things, then we should see that behavior reflected in a growing share of citations to the most cited works. We measured semantic diversity based on paper and patent titles; values correspond to the ratio of the standard deviation to the mean pairwise cosine similarity (i.e., the coefficient of variation) among the titles of the 1% most cited papers and patents by field and year. We chose to focus on titles (rather than abstracts) for consistency with our other analyses in the manuscript (which are mostly based on titles) and because titles are available for WoS papers over the entire study window (whereas abstracts are only reliably available beginning in the early 1990s). To enable semantic comparisons, titles were vectorized using pretrained word embeddings.

The results of these analyses are shown in an updated Figure 4A and 4D (inset plots) and discussed in the body of the main manuscript (p. 7 lines 156-158), with additional details in the caption for Figure 4 (p. 12). We also show larger versions of the plots below, for convenience. Values of both measures (i.e., share of citations to the top 1% most cited and semantic diversity of the top 1% most cited) are computed within field and year, and are subsequently averaged across fields for plotting. Consistent with what we might anticipate, we find that the decline in diversity of work cited is accompanied by an increase in the share of citations to the 1% most highly cited papers and patents (Figures 4A1 and 4D1), which are also decreasing in semantic diversity (Figures 4A2 and 4D2). Put a different way, over time, scientists and inventors are increasingly citing the same prior work, and that prior work is becoming more similar.



R2C6: *The other findings feels someone oddly interpreted: is self-citation an indication that scientists and inventors are more focused on their own work, or there is more of that work to be cited (as scientists age). Moreover, are aging citations an indication that scientists "struggle to keep up" or that the stock of relevant work is increasing (see my first point above) and/or that aging scientists are actively fighting against change? I do not doubt that the three measured factors are causally related to disruption, and co-related to one another, but the interpretation seems to miss within-field and within-person forces that do not diminish their importance, but change their interpretation and neutralize the associated policy implication (write less, read more—P. 7—makes less sense unless the writing less prescription is redefined to be not over historical time but over the research career; and out of field). I note that I did find the observation of a conservation of disruptive papers and patents a very interesting finding, mic-dropped at the end, but would have loved for this to have been detailed/demonstrated/interpreted further.*

Thank you for your careful reading of our regression models. Your comment includes several points, which we address in turn.

First, we agree with your concern that in the prior draft, the measures of self-citation and citation age in our regression were open to multiple interpretations, which in turn clouded the policy implications of our analysis. We addressed this concern in two ways: (1) we consulted the literature on self citation and citation age, and (2), we added controls for the average (career) age of scientists and inventors on paper/patent teams and for their career publication/patenting activity.

Beginning with (1) we systematically searched for prior literature on self citation in the Science of Science. Prior works have interpreted self citation as an indication of the extent to which one builds on his or her own pre-existing research stream (e.g., Bonzi and Snyder 1991; Fowler and Aksnes 2007; King et al., 2017). In addition, we also looked for prior work on citation age. Existing research has used citations to older work as a proxy for reliance on dated knowledge and a simultaneous failure to keep up with the latest discoveries (e.g., Merton 1961; Wang et al. 2013, Mukherjee et al 2017). Overall, while we agree that there are multiple factors likely driving both the increase in self citation and citation age, we find that generally, prior work has understood these in ways that are similar to those we propose in the paper (i.e., indicating narrower search and use of prior work). Thank you for pushing us to think carefully about how we interpret these measures. We now include references to these sets of relevant prior works on the two measures when we interpret them (p. 7 lines 159-164).

Turning to (2), we also sought to allow for a clearer interpretation of our regression models by adding new control variables to the models. First, your comment notes that rather than indicating a greater focus of scientists and inventors on their own work, self-citation may reflect the amount of prior work published by scientists and inventors (and therefore available for self citation). To account for this possibility, we added a control for the *Mean number of prior works producedby team members* to Models 6 (papers) and 12 (patents) of Extended Data Table 2; "prior works" are defined as the cumulative papers or patents, respectively, published by members of the focal team at any time prior to the focal paper or patent. Second, you raise an important point that our measures of citation age are difficult to interpret because they could be confounded with the age of the authors/inventors. To account for this possibility, we now include a control (also in Models 6 and 12 of Extended Data Table 2) for the *Mean age of teammembers*. Following prior work (e.g., Cui et al., 2022), we measure age as "career age", defined as the difference between the publication year of the focal paper or patent and the first year in which each author or inventor published a paper or patent. We find that coefficient estimates for our measures of self-citation and citation age are substantively and statistically similar after including the new controls.

To complement these results, in unreported analyses (available upon request), we also ran additional models that attempted to better account for the stock of relevant prior work at the field level. First, in one series of models, we included additional control variables for the total number of unique papers/patents cited by papers/patents in the focal subfield and year. Second, we ran models that added, in addition to fixed effects for subfield and year (which are included in our main models), fixed effects for their *interaction*. These latter models are particularly conservative, as they account for all unmeasured factors that vary by field and year (including, e.g., the volume of relevant prior work). In both of these additional analyses, we continue to find consistent results.

In summary, to the best of our knowledge, our interpretation of our measures of self citation and citation age are consistent with prior work in Science of Science. Moreover, we also believe that our finding of consistent results after the addition of the new control variables discussed above helps to give additional confidence to our interpretation. Nevertheless, despite these efforts, we also acknowledge that given the observational nature of our data, we are unable to definitively rule out all alternative explanations.

Second, we appreciate your positive comments on our findings regarding the stability of disruptive papers and patents (as shown in Extended Data Figure 5), which we also thought were interesting. We now mention this notion of "conservation" in our discussion of those findings (see p. 5). Moreover, following your suggestion, we expanded this analysis by adding new inset plots to Extended Data Figure 5 that show the composition of the most disruptive works by field, over time. Remarkably, we find that the stability in the absolute number of highly disruptive papers and patents holds despite considerable churn in the underlying fields responsible for producing those works. While we can only speculate, this finding may indeed suggest that there is something of a "global" carrying capacity (a concept we decided to further explore based on your point in **R2C9**) in the number of disruptive papers and patents that can be produced by science and technology at any given time, but that "locally" (i.e., within fields), that capacity is more flexible, at least in the short term, perhaps reflecting the shifting interests of funders and scientists and the "ripeness" of scientific and technological knowledge for discovery and invention.

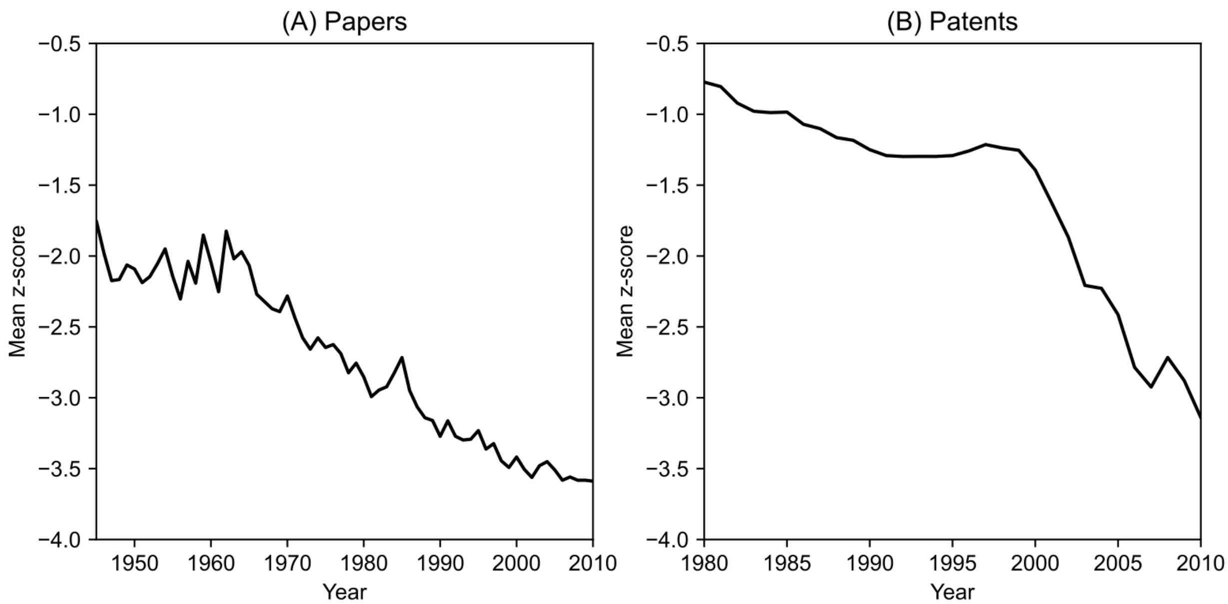
R2C7: *I did not fully understand or was persuaded by the author's demonstration that "this trend is unlikely driven by change in citation practices or the quality of published work." The first is supposedly demonstrated by rewiring citations to preserve network structure and age and showed a similar CD actual/rewired ratio (hovering, fluctuating around .75, but dropping for physical science and rising above 1 for electrical and electonic patents). But the change if the change in citation patterns has to do with an increase in relevant past work (which it does) or the change in network structure (e.g., entropy, which it also does) then what other style of citation change was this supposed to exclude? The quality of published work shows that papers in top outlets manifest the same as all other work, which suggests that this is a secular change, which may be about decreasing innovation, but may also be about prior relevance.*

Response: Thank you for your careful attention to the underlying mechanisms and for drawing our attention to some potential weaknesses in our analysis. Your comment has several parts, which we address in turn.

First, we agree that in the previous version of the manuscript, we had not done enough to address whether our results may be driven primarily by changes in citation practices, and in particular the increasing relevance of prior work. In the revised manuscript, we have added many new analyses that aim to account for this important alternative explanation—including creating new, normalized indicators, regression adjustment, and simulation analysis—which are collected in a new section of the paper entitled, "Robustness to changes in publication, citation, authorship practices." We describe these analyses in greater detail in response to your comment #1 **(R2C1)**. We recognize that the declines we observe in the CD_5 index are likely driven by multiple factors, and we agree with your point that the changing volume and relevance of prior science and technology is likely an important contributor. However, based on our assessments—including multiple approaches for accounting for alternative explanations, and finding of consistent patterns with alternative indicators (e.g., novelty, language, and word use)—we also believe that the evidence strongly suggests that the observed decline in the CD_5 index is not driven primarily by changes in the volume and relevance of prior science and technology.

Second, we also appreciate you drawing our attention to the lack of clarity in our simulation analysis. To that end, we have completely revised both the simulation analysis and its presentation, with the goal of making the rationale, approach, and interpretation much clearer (see the subsection, "Verification using simulation" in the section, "Robustness to changes in publication, citation, authorship practice," p. 22-23 lines 458-502, for more detail). Among the changes, we now follow an approach closer to that of Uzzi et al., (2013)'s paper on "atypical combinations," specifically presenting z-scores that compare the observed values of the CD index to those from the rewired networks, which we believe may be more familiar to readers without expertise in network science. In particular, whereas in the prior version of the manuscript, we had compared the CD_5 index values from the observed network to those from a single rewired network, in the revised paper, we now compare the observed values to those from 10 rewired copies each for papers and patents (following Uzzi et al., 2013), which allows us to compute a z-score for each paper/patent.

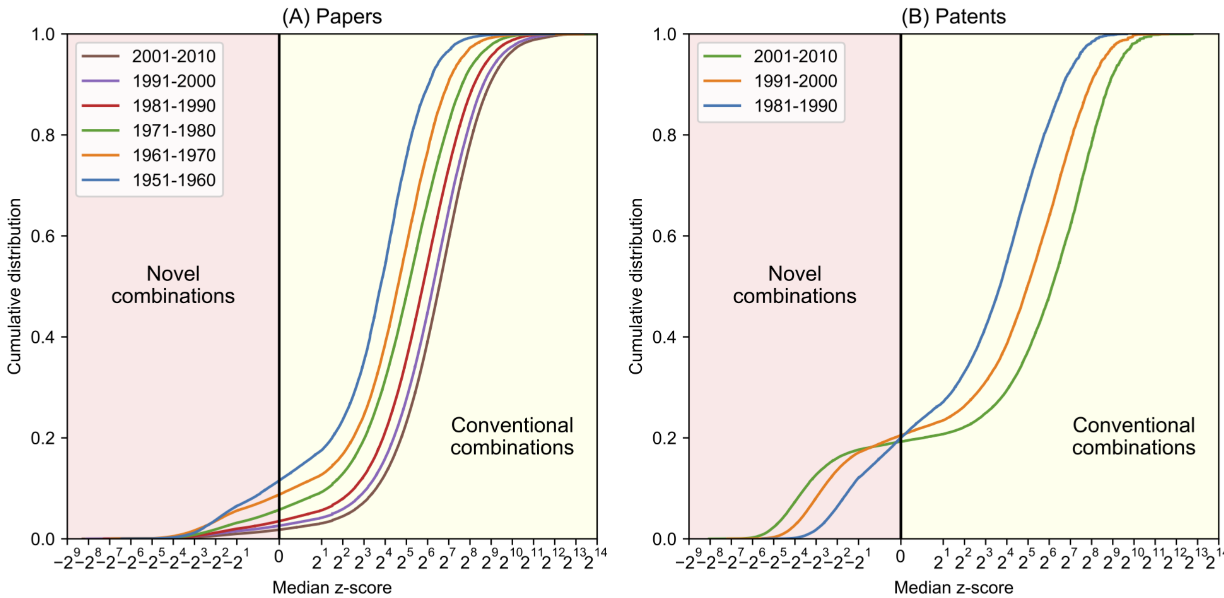
To clarify a bit more substantively, as you note in your comment, the motivation for our simulation analysis was the concern that the observed declines in the CD_5 index could be driven at least in part by changes in citation practices. Over time, papers and patents are citing more prior work, particularly as the volume and relevance of prior science and technology available for citing increases. Consequently, there are also more opportunities for future papers or patents to cite the references of a focal work (both because the focal work is likely to have more references that are at risk of being cited, and because future work is making more citations to prior work), which would in turn result in declining values of the CD_5 index, even by chance alone. Our simulation analysis helps to account for this possibility. Specifically, in that analysis, we rewire the observed citation networks by randomly swapping edges such that for each paper and patent, *the number of citations made, the number of citations received, and the ages of the citing and cited works are identical in the observed and rewired networks.* Consequently, if the decline in the CD_5 index is driven primarily by changes in the these structural factors, then we should see a decline over time in the CD_5 index values obtained from the rewired networks that is comparable to the decline in the observed networks (because the number of citations made, the number of citations received, and the ages of the citing and cited works are identical in the observed and rewired networks). Consistent with the intuition sketched out above, we do indeed find (in unreported analyses, available upon request) that the values of the CD_5 index in the rewired networks declines over time, suggesting that some component of the decline is attributable to structural changes in the citation networks. However, as evidenced by plots of the z-scores comparing the average CD_5 index values from the observed citation networks to those in the rewired networks (see p. 42 Extended Data Figure 12, also shown below for your convenience), we also find that the decline in average CD_5 index based on the observe networks outpaces that based on the rewired networks, suggesting that structural changes in the citation networks alone do not account for the decline.



Finally, your comment mentions our analysis in Extended Data Figure 7 (p. 37), where we plot the CD_5 index values for papers associated with the awarding of a Nobel Prize and papers published in widely recognized high-quality journals (i.e., Nature, Science, and PNAS). We agree that these descriptive plots cannot alone rule out whether the observed decline in the CD_5 index is due to the changing relevance of past work or changing innovation (or both). Instead, our goal with these plots was to determine whether the decline was also observable for work that is widely recognized to be "high quality." As your comment notes, the patterns we find in "high quality" work are similar to those we see across the broader population of papers, which is indicative of a secular decline. Based on our other analyses in the paper, we believe that the decline is reflective of a decline in disruption, however the plots in Extended Data Figure 7 alone cannot establish the cause of the decline. We apologize for any confusion our presentation may have caused.

R2C8: *What was notably absent here was discussion/measurement of one of the many operationalizations of combinatorial novelty (Uzzi et al. 2013; Sorenson, Rivkin, and Fleming, n.d.) over time, despite efforts to measure the described performance of novel things (e.g., the semantic categories of "creation" and "discovery"). Obviously, with a greater number of things studied, opportunities for combinatorial novelty increase over time and so this is less likely to drop in the same way that disruption does, but we should know this based on its importance to advance and its clear relation to creation, discovery and invention in both science and technological domains.*

Response: Thank you for this suggestion. Based on your feedback (which was also echoed in comments from Reviewer #1 **[R1C4]**), we computed Uzzi et al.'s (2013) measure of atypical combinations for both papers and patents. For papers, our data (Web of Science) are identical to Uzzi et al.'s and we were able to follow their approach exactly. For patents (which were not studied in Uzzi et al.), we adapted the Uzzi et al. approach to examine pairings among the primary USPC classes of the cited patents (rather than journals). We present the results of this analysis following the approach of Uzzi et al., who plotted the cumulative distribution function of their novelty measure, with separate lines for subsets of interest. In our case, we plot separate lines for papers and patents published in each decade of our study window. The results are shown below and also included in the revised manuscript in (p. 34 Extended Data Figure 4).



Overall, we find a rightward shift in the cumulative distributions over time, suggesting that for both papers and patents, combinations are more conventional than would be expected by chance, consistent with what we would anticipate based on our results using the CD index. Interestingly, for patents, there is also a smaller shift in the opposite direction on the left side of the distribution, suggesting that novel patents in recent decades are somewhat *more* novel than novel patents in earlier decades. Still, the bulk of the distribution is moving to the right, indicating greater conventionality.

In addition to showing changes in combinatorial novelty using Uzzi et al.'s (2013), we also added a new analysis where we examine the novelty of word pairs appearing in the titles of papers and patents over time (see Figure 3 for the results; our methodological approach is described in Section 1 of the Supplementary Information [p. 47-48 lines 678-683]). Consistent with our findings using Uzzi et al.'s (2013) measure, we observe a persistent decline in new word pairs over time.

R2C9: *Still and all, the paper boldly documents a big trend (caveats as stated above) and is clearly written (despite incompleteness in the math of their measure and the procedure for their survey, which could trivially be fixed). A mild, final concern relates to the theme of slowing science, which has become something of a commonplace in science, based on several studies they cite (Bloom et al. 2020; Jones 2009; Chu and Evans 2021; Jones and Weinberg 2011; Wu, Wang, and Evans 2019; Milojević 2015), but also editorialized more broadly (e.g., "Science is Getting Less Bang for Its Buck" by Collison and Nielsen in the Atlantic). What exactly do the authors contribute to this narrative. They show that there is more cited work in the past perceived as relevant, more uneven/unequal attention to that work ("diversity"), and fewer works seen as framing new directions with the policy prescription to write less and read more. This work contributes to the narrative, but it doesn't change it. To be clear, however, it is an important narrative—maybe the most important for the Science of Science and Innovation—and one worth understanding better. Moreover, there were hints of things that do or could contribute in new ways, like the seeming conservation of disruption mentioned in the last paragraph that could suggest something of a carrying capacity in science.*

Response: Thank you for your kind words and for highlighting the potential for our analysis to make an important contribution. We acknowledge that our paper is not the first to suggest that science is slowing, and we have drawn attention to relevant prior work as appropriate throughout the paper, including the papers you mention in your comment. We also believe that we make several distinct contributions over and beyond what has been demonstrated by prior work, and we hope these contributions come through more clearly in the revised manuscript.

First, to the best of our knowledge, our study is the first to systematically document declining disruption in science and technology across a broad range of different fields, using a consistent, validated, and state of the art bibliometric measure. More specifically, while our study is not the first to document slowing progress, existing research has tended to do so through studies of particular fields, using disparate and domain-specific metrics. For example, in economics, much of the evidence for the slowdown comes from the semiconductor industry, where the number of transistors on an integrated circuit serves as a common measure of progress (e.g., Bloom et al., 2020; Cowen and Southwood, 2019). Studies have also focused on the pharmaceutical industry, where a key metric of interest is the number of new drug approvals (e.g., Pammoli et al., 2011). Recent work has also shown evidence of slowing progress in artificial intelligence, where algorithmic performance is a key benchmark (Besiroglu, 2020). Researchers working in the Science of Science have also provided important evidence of slowing progress, but existing studies tend to focus on specific fields, and/or provide such evidence indirectly and as part of a larger study. As an illustration, Chu and Evans (2021) find that "canonical progress" is slower in large fields of science, as indicated by an inverse relationship between the number of papers published each year and churn in the most cited works in a field. Although one may imagine that changes in "canonical progress" are related to changes in disruption, the authors do not test that relationship explicitly, nor do they demonstrate changes in disruption over time. While all of this work has made important contributions, the tendency of existing research to focus on particular fields, using disparate and domain-specific metrics has made it difficult to know whether the changes are happening at similar rates across areas of science and technology, thereby leaving an important gap in understanding—one we believe we fill with our study.

Second, we also propose and find support for a novel mechanism—changes in the use of prior knowledge—that helps account for the decline in disruptive science and technology. Moreover, our findings on that mechanism make an important contribution not only by identifying a factor leading to declining disruption, but also, perhaps even more critically, by helping **to reconcile an increasingly difficult theoretical tension in the Science of Science around theories of recombination and the endogenous growth of knowledge.** In particular, one of the most foundational observations in the Science of Science is that innovation is a process of recombination, wherein breakthroughs often result from bringing together existing components (e.g., ideas, concepts, materials) in novel ways. As science and technology have grown and continue to grow over the past decades, there would seemingly be more "existing components" available for recombination than ever before, which intuitively would suggest a speeding up, rather than a slowing down, of innovative activity. Within this context, our demonstration that over time, scientists and inventors are using a narrower scope of existing work and, moreover, that using a narrower scope of existing work is associated with less disruption, offers one way of reconciling this important theoretical tension.

Finally, we also believe our paper makes a contribution by demonstrating that the decline in disruptive science and technology is robust to many alternative explanations. For example, by replicating our analysis across WoS, PatentsView, JSTOR, APS, Microsoft Academic Graph, and PubMed, we demonstrate that our finding is not an artifact of the data source used. We also show that the decline is not an artifact of our particular mathematical operationalization of disruptiveness, as it is also observable on other recently proposed indicators. Moreover, we also show evidence of the decline on non-bibliometric indicators, including through large-scale text analyses. Finally, thanks to helpful input from your review and other members of the review team, we have also conducted many new analyses that help add confidence that our results are not an artifact of changes in publication or citation practices over time.

Overall, we would like to thank you for your kind words on our paper, and for pushing us to think more carefully about our contribution, which we hope comes through more clearly for the reader in the updated manuscript.

R2C10: *Figure 1. Fails to describe the function in C: no definition of f_it, b_it or how the measure works despite its (uninterpretable) illustration. Also, it describes CD_t=0 as its median value, but this, obviously hinges on the distribution of CD_t values and relates to my earlier main point about a lack of care regarding CD distributions that are neither described nor graphed and illustrated (except for here, with the almost assuredly false characterization that 0 is median).*

Response: Thank you for bringing these issues to our attention. As we noted in response to your comment **R2C2**, we revised Figure 1 to include explicit definitions of all components of the CD index and to correct the error regarding the median value. We also added clarification on the interpretation of the strip plots in Panel B of Figure 1. Finally, as described in more detail above, we also added a new figure, Extended Data Figure 1 (p. 31), which provides more detail on the distribution of the measure for papers and patents.

R2C11: *Extended Data Fig. 1A is broken: the measurements prior to 1992 to appear to missing...maybe from a rendering error (but is this supposed to be the same as Figure 3A?...I think so, but then why do 3B and ED1B not appear to line-up?)*

Response: We apologize for the confusion. Extended Data Figure 3A is based on the text of papers' *abstracts*, which are only reliably available in Web of Science beginning in 1992 (as indicated in the caption of of the figure [p. 33]; see Section 1 of the Supplementary Information for more discussion [p. 46 lines 656-662]). Figure 3A (p. 11) is based on the text of papers' *titles,* which is available on Web of Science from 1945.

R2C12: *SI Line 593: "A small number of words APPEAREDin disruptive"*

Response: Thank you for bringing this to our attention. We have made the correction as suggested.

R2C13: *SI heading 3 "Classification of disruptiveand consolidating words". The authors mentioned an "All Our Ideas" survey and the three people who filled it out,*

Response: Thank you for this comment. Unfortunately, in our copy of the reviews, the full text of this comment was cut off mid sentence. However, based on the text that was included, we believe your comment was regarding the "All Our Ideas" survey, which was used to distinguish "consolidating" and "disruptive" words in the prior version of Figure 2. As we noted in response to your previous comment **(R2C4)**, based on feedback from you and similar comments from other members of the review team, we have omitted this analysis from the paper. We hope that doing so addresses your concern.

Response to Reviewer #3

(*Original Comments in Italics* / Author Responses in Standard Typeface)

R3 General: *This is a novel contribution to an important conversation about whether the research enterprise is generating less progress despite the strong growth in production of papers. The authors focus on a metric measuring the extent to which papers do not also cite the references of the papers they reference. The logic is that if a paper is a breakthrough, it shifts the conversation into new directions and papers that build on the breakthrough paper no longer need to pay attention to what the authors of the breakthrough paper were attending to. On the other hand, if a paper adds to an ongoing conversation, then papers that build on it will be motivated by the same work and will therefore cite the same work. The rationale makes sense to me.*

Other papers have opened up the possibility that progress is slowing, but as the authors state the basis for those arguments has been field specific. This paper tackles the question across the full research domain. The OECD is concerned about the problem. They are producing an edited book on the topic and whether AI can offer solutions based on summer conference. I take this as evidence that this is a high level problem of international concern.

Response: Thank you for your kind words on the contribution of our research and for your helpful suggestions for improving the manuscript. We also appreciate you highlighting the connection between our research and the OECD policy agenda, which we now highlight in the introduction of the paper (p. 2 lines 19-21). We have worked hard to try to fully address all of your concerns, and hope that you find this revised version to be a significant improvement to our previous draft.

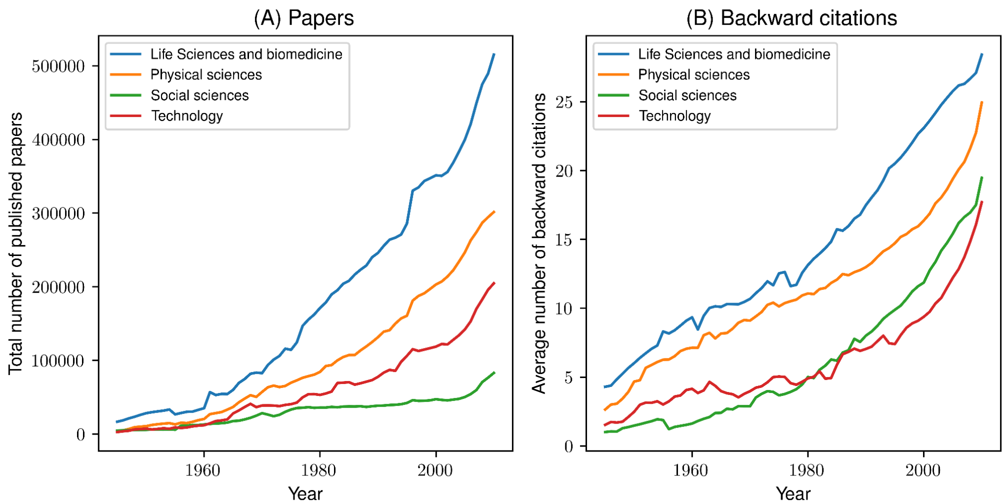
R3C1: *A more nuanced reading is warranted of Figure2: Decline of disruptive science and technology. This is the display of the main result, the declining metric of disruptiveness. The paper figure begins in 1945 and the interpretation in the text reports decline since 1945. But is it really surprising that the world today differs from the world in 1947? Not so much. If we look at the situation in papers since 1980, physical sciences has been stable and life sciences pretty much so except for a step change (which I never trust in such data) in about 1997; WoS may have changed its coverage in 1997. Social sciences shows the most dramatic and continuous drop since 1980. The disruptiveness in patents has been stable for about the last quarter century, i.e. since about 2000.*

Response: Thank you for your careful attention to the presentation of our results. Your comment includes several points, which we address in turn.

First, we agree that in the prior manuscript, our discussion of Figure 2 was too focused on the overall decline in disruptive science and technology and did not present a sufficiently nuanced description of the trends. We have updated our discussion of Figure 2 (see p. 4 lines 62-75) to include a more detailed presentation of the trends, with a focus on highlighting the periods of the greatest decline and differences across fields.

Second, as your comment suggests, science has changed dramatically since 1945, which raises concerns about the validity of our comparisons of the CD_5 index over long periods of time. This concern was echoed in your comment #5 **(R3C5)** and in feedback from the other reviewers (e.g., **R1C2**, **R2C1**). In the revised manuscript, we have undertaken extensive new analyses with the goal of ensuring our comparisons across time are valid, including by implementing your suggestions in **R3C5** below. For additional details, please see our response to your comment #5.

Finally, your comment mentions changes in WoS coverage as a potential explanation for the step change in Life Sciences and Biomedicine around 1997, which is a concern we shared. Based on your feedback, we conducted several additional analyses to more systematically examine whether changes in coverage may account for any of the patterns observed in Figure 2. First, we examined the WoS documentation along with several published papers on coverage changes in WoS and did not find evidence of changes that would clearly account for the patterns observed around 1997 (Adriaanse and Rensleigh, 2013; Clarivate, 2022; de Winter et al., 2013; Fortunato et al., 2018; Singh et al., 2021). Second, we examined the data to look for evidence of changes in coverage or structure. For example, the journals archived by WoS may have shifted during the period around the fluctuations in question, leading to changes in the structure of the citation data, which may account for the step change. If this were the case, we would expect to see proportional shifts in the number of papers covered by Web of Science and/or changes in the count of backward citations made by papers during the periods of fluctuation. Based on these considerations, we plotted the number of papers and the average number of citations made by papers in WoS (shown below).



Overall, while we see some evidence of an increase in the number of papers in the Life Sciences and Biomedicine in the late 1990s, we do not see a corresponding change in the average number of backward citations, and moreover, the increase does not deviate substantially from the overall trend. Based on these assessments, we do not believe that changes in WoS coverage are primarily responsible for the various local peaks or the overall trends we report in Figure 2.

R3C2: *I recommend dropping "We verify the decline in disruptiveness over time through regression models, which show that publication year has a statistically significant and negative relationship with CD5 for papers (p < 0.001) and patents(p<0.001)." The drop visible in the graph does not need verifying. Instead, more subtlety in interpretation is required.*

Response: Thank you for your suggestion. We have now removed this line from the text in the revised manuscript (along with the associated section of the Methods). Instead, we followed your advice (particularly from the previous comment) to provide a more nuanced discussion of the trends in disruptiveness reported in Figure 1 (see especially the updated text on p. 4 lines 62-68). In addition, as we describe below, in response to your point **R3C5**, we conducted a series of regression analyses aimed at verifying the decline in a way that is more robust to changes in publication, citation, and authorship practices.

R3C3: *I would like more clarity on some aspects of method. 1 On alternative samples. As written the text implies analysis was run on JSTOR and APS. But that is impossible. The metric concerns citing and referencing. JSTOR and APS do not index references and citations. Was the analysis performed on records indexed in WoS that are also indexed in JSTOR and APS, and PUBMD? In fact, was the Microsoft Academic Graph analysis also done by extracting records from MAG and then looking at the metrics calculated in WoS? That would require less effort so I assume that was how it was done rather than rebuilding the system in MAG. The text should describe exactly what was done, including how many items from each source could not be found in WoS.*

Response: Thank you for your careful attention to our data and methods. For both the JSTOR and APS data, we actually were able to obtain citation data (i.e., lists of citing and cited papers) directly from their respective publishers (i.e., we did not need to link either database to WoS or MAG). For reference, the APS data we used is described here (see the heading, "Citing article pairs"): https://journals.aps.org/datasets. In the revised manuscript, under the subheading "Alternative samples" in the "Methods" section, we now provide additional details on where we obtained each of the supplemental datasets (p. 20 lines 366-375).

R3C4: *2 On the Monte Carlo simulation. The origin of the metric lies in network analysis. As network analysts, the authors find convincing a network analysis technique of comparing the data to a randomly rearranged network. However, not being a network analyst, I don't see the rationale here. It would seem that all that is achieved by comparing it to a random graph is establishing that it is not random. We know that the citation graph is not random. We don't need a Monte Carlo simulation to establish that.*

Response: Thank you for drawing this lack of clarity to our attention. We agree that in the previous draft, the rationale for the Monte Carlo simulation was not well articulated. As we describe in more detail in response to your next comment, we have revised that analysis with the goal of making the motivation, approach, and interpretation much clearer, especially for non-network readers.

R3C5: *I would like to see that dropped and replaced with a regression that establishes the extent to which the change in the metric over time is explained by change in number of papers published, number of references per paper, and number of authors per paper. The very last item in the supplementary information does part of this. So I am saying move it to the main text and include trends in number of references per paper and authors per paper, both of which could be plausibly related to trends in the metric. This is not the same as extended table 2 which seeks to explain a paper's metric using characteristics of that paper.*

Response: Thank you for drawing our attention to this important concern. Indeed, the possibility that our results may be accounted for by changes in publication, citation, or related practices was echoed by other members of the review team (e.g., **R1C2**, **R2C1**). For that reason, we were especially grateful for your concrete suggestions on how to address the issue, which we have implemented in the revised manuscript.

Specifically, following your suggestion, we now report the results of regression models (p. 51 Supplementary Table 2) predicting the CD_5 index of papers (Models 1-4) and patents (Models 5-8), with indicator variables included for each year of our study window (for details, see p. 21-22 lines 434-457). The most conservative estimations (Models 4 and 8 for papers and patents, respectively) include subfield fixed effects and control variables for several field × year level—*Number of new papers/patents*, *Mean number of papers/patentscited*, *Mean number of authors/inventors per paper*—and paper/patent-level—*Number of papers/patents cited*—characteristics, thereby enabling more robust comparisons in patterns of disruptive science and technology over the long time period spanned by our study. Across models, we find that the coefficients on the year indicators are statistically significant and negative, and growing in magnitude over time (see p. 41 Extended Data Figure 11 for a visualization of the model predictions), consistent with the patterns we reported based on unadjusted values on the CD_5 index in Figure 2. These results are now discussed in the main text (p. 6 lines 133-137) and the caption to Figure 2 (where we also refer the reader to a more extended methodological overview and presentation of the results in the subsection of the "Methods" section entitled, "Robustness to changes in publication, citation, authorship practice" (p. 21-22 lines 434-457).

In addition to your suggested regression-based approach, we also conducted several other analyses to further ensure our findings on the decline of disruptive science and technology are not the result of changes in publication, citation, or related practices.

1. First, following suggestions from Reviewer #1 **(R1C2)**, we developed several normalized versions of the CD index that aim to facilitate comparisons over time (p. 40 Extended Data Figure 10), particularly by adjusting for the increasing tendency for papers and patents to cite prior work. Consistent with our findings in the main text, we observe a decline in disruption on these normalized indicators. Additionally, to complement our two normalization approaches, we also computed an alternative, previously developed variation on the CD index that is less susceptible to changes in citation practices. Specifically, as discussed in the manuscript under "Alternative bibliometric measures" (p. 20 lines 376-388), Bornmann and colleagues (2020) proposed a novel disruption

indicator, $DI_l^{no\ k}$, that by design does not include the N_k term, and that therefore may enable more robust comparisons across time periods. As we show in Extended Data Figure 9 (p. 39), we find that the average value of $DI_l^{no\ k}$ declines over time, similar to what we see using the original CD index. We hope these results add reassurance that our findings are not driven by changes in citation practices.

2. Second, we agree that the simulation analysis included in the previous version of the paper was confusing and did not sufficiently address concerns that our findings may be an artifact of the data. In revising the paper, we had planned to implement your suggestion of dropping that analysis from the paper. However, we also thought that with some changes in the approach, and principally with a clearer explanation of the motivation and interpretation, the analysis might serve as a helpful complement to the regression and normalization approaches just described, as it uses a very different methodology to reach similar conclusions (i.e., that our results are not primarily driven by changes in citation or publication practices). To that end, we revised the simulation analysis and its presentation, with the goal of making the rationale, approach, and interpretation much clearer (see the section, "Robustness to changes in publication, citation, authorship practice" for more detail). Among the changes, we now follow an approach much closer to that of Uzzi et al., (2013)'s paper on "atypical combinations," specifically presenting z-scores that compare the observed values of the CD index to those from the rewired networks, which we believe may be more familiar to non-network readers. Of course, we recognize that even despite our revisions, the simulation analysis may feel less clear or compelling than our other approaches for adjusting for changes in publication and citation practices, in which case we would be happy to completely drop it from the paper. We would really appreciate the opportunity to have your feedback on this point.

R3C6: *On details of presentation. I was somewhat frustratedtrying to piece together the material spread across manuscript with references, methods with references, extended data and supplementary information. However, the fault lies not with the authors but with* Nature *not being able to afford social science papers the room they need to explain their material thoroughly. Anyway, some details need cleaning up.*

Response: We apologize for the scattered nature of the figures and tables. In the revised manuscript, we have done our best to make the presentation as clear as possible.

R3C7: *Extended figure 1 is not mentioned in the extended text. It feels like it is hanging out there by itself. It should be mentioned in the extended text.*

Response: Thank you for this suggestion. In the manuscript, we ordered the figures in the Extended data section based on the order in which they are referred to in the main text. For example, Extended Data Figure 3 ("Extended Data Figure 1" in the previous version of the manuscript) is mentioned in the main text on p. 4 lines 84-87. (We would also like to highlight that we provide more information about Extended Data Figure 3 in the first section of the Supplementary Information, p. 46-47 lines 644-683).We recognize that the many tables, figures, and appendices result in a complex web of cross references throughout the text, which can be difficult to follow. In the revised manuscript, we have tried to make the presentation as simple and intuitive as possible given the formatting requirements of *Nature*. We hope that this clarification addresses your concern.

R3C8: *Please use the non-jargon explanation of lexicaldiversity – unique/total words because without the explanation (provided in the text but not the extended text) the jargon will be obscure to many readers*

Response: Thank you for this suggestion. We have now removed the more esoteric term "lexical diversity" in favor of "unique/total words" throughout the manuscript.

R3C9: *Extended figure 2 – We know papers have increased a lot. Do not plot them here (blue line) because it so compresses the y axis that we cannot effectively judge the claims about the papers of interest, those with CD>0.25.*

Response: Thank you for this suggestion. To clarify, the blue line represents the count of papers with CD index values in the range (0.00, 0.25] (i.e., not total papers); papers with CD index values of 0.00 or lower are excluded from the figure. We have updated the figure caption to more clearly describe the sample.

R3C10: *Supplementary information, lexical diversity – Was uniqueness defined within a field or across the whole dataset? How do we know that even obsolete words are occasionally used? Is there a reference? Your analysis? Obsolete words are used in titles and abstracts? The second sentence in this section talks about increases in disruption, but the paper searches for decreases in disruption. Change "increases" to "decreases".*

Response: Thank you for your careful review of our supplementary material and for drawing our attention to several areas in need of clarification. Your comment includes several points, which we respond to in turn.

1. Regarding the measurement of lexical diversity, you are correct that we defined uniqueness (and total token count) within fields, not across the whole dataset. In the revised manuscript, we updated the caption for Extended Data Figure 3 (p. 33) to better reflect this point. We also include more details in Section 1 of the Supplementary Information (see especially the paragraph on p. 46-67, lines 670-677).
2. Regarding the use of "obsolete" words, in the updated manuscript, we now provide support for those claims, which comes primarily from prior work on linguistic change (e.g., Michel et al., 2011), but is also evident in our data. For example, in the WoS data, we find that words like "phlogiston," "aether," "preformism," and "lamarkian"[2] continued to appear in the titles of scientific papers (though in much lower frequencies) well after their associated scientific theories were "superseded."
3. Finally, regarding the phrasing of "increases" versus "decreases" in disruption, we updated Section 1 of the Supplementary Information to use "decreases," consistent with the rest of the manuscript.

R3C11: *Supplementary table 1 – The table displays verbs. The table note is all about nouns. Replace current note with explanation of the colors of the words and include black in that explanation. Also explain how the numbers were calculated.*

Response: Thank you for your careful reading of our tables and for drawing that inconsistency to our attention. Reviewers #1 (e.g., **R1C5**) and #2 (e.g., **R2C4**) also highlighted limitations of our analysis of changes in word use over time, particularly with respect to our approach for identifying "disruptive" and "consolidating" words. In the revised manuscript, we significantly reworked our analyses of word use, including by dropping several analyses that seemed to add more confusion than clarity. Supplementary Table 1 ended up being one of those that we dropped.

R3C12: *On English usage, much use is made of "utilize". Use is preferable.*

Response: Thank you for this suggestion. We agree that "use" is clearer, and in the revised manuscript, we replaced instances of "utilize", "utilization", and so forth with alternative terms.

[2] We identified these terms from the Wikipedia page on "Superseded theories in science": https://en.wikipedia.org/wiki/Superseded_theories_in_science

REFERENCES

Adriaanse, L. S. and Rensleigh, C. (2013), 'Web of science, scopus and google scholar: A content comprehensiveness comparison', *The Electronic Library* .

Appel, T. A. (2000), *Shaping Biology: The National Science Foundation and American Biological Research, 1945-1975*, JHU Press.

Besiroglu, Tamay. (2020), *Are models getting harder to find*? Masters Thesis, University of Cambridge.

Bloom, N., Jones, C. I., Van Reenen, J. and Webb, M. (2020), 'Are ideas getting harder to find?', *American Economic Review* 110(4), 1104–44.

Bonzi, S. and Snyder, H. (1991), 'Motivations for citation: A comparison of self citation and citation to others', *Scientometrics* 21(2), 245–254.

Bornmann, L. (2020), 'How can citation impact in bibliometrics be normalized? A new approach combining citing-side normalization and citation percentiles', *Quantitative Science Studies* 1(4), 1553–1569.

Bornmann, L., Devarakonda, S., Tekles, A. and Chacko, G. (2020), 'Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers', *Quantitative Science Studies* 1(3), 1242–1259.

Bornmann, L. and Marx, W. (2015), 'Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts?', *Journal of Informetrics* 9(2), 408–418.

Bornmann, L. and Tekles, A. (2019), 'Disruption index depends on length of citation window', *El profesional de la informacion (EPI)* 28(2).

Burt, C. (1955), 'The meaning and assessment of intelligence', *Eugenics Review* 47(2), 81.

Certttti, S., Longbxni, E. and Pinciholi, F. (1979), 'A computer-oriented approach to the comprehensive organization of information in hand-compiled medical records', *Methods of Information in Medicine* 18(03), 138–145.

Chu, J. S. G. and Evans, J. A. (2021), 'Slowed canonical progress in large fields of science', *Proceedings of the National Academy of Sciences* 118(41).

Clarivate (2021), 'Web of Science Coverage Details', https://clarivate.libguides.com/librarianresources.

Collison, P. and Nielsen, M. (2018), 'Science is getting less bang for its buck', *Atlantic* .

Cowen, T. (2020), 'The nobel prize isn't what it used to be', *Bloomberg*.

Cowen, T. and Southwood, B. (2019), 'Is the rate of scientific progress slowing down'.

Cui, H., Wu, L. and Evans, J. A. (2022), 'Aging scientists and slowed advance', *arXiv preprint arXiv:2202.04044*.

De Winter, J. C., Zadpoor, A. A. and Dodou, D. (2014), 'The expansion of google scholar versus web of science: a longitudinal study', *Scientometrics* 98(2), 1547–1565.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B. et al. (2018), 'Science of science', *Science* 359(6379), eaao0185.

Fowler, J. and Aksnes, D. (2007), 'Does self-citation pay?', *Scientometrics* 72(3), 427–437.

Jones, B. F., Wuchty, S. and Uzzi, B. (2008), 'Multi-university research teams: Shifting impact, geography, and stratification in science', *Science* 322(5905), 1259–1262.

King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J. and West, J. D. (2017), 'Men set their own cites high: Gender and self-citation across fields and over time', *Socius* 3, 2378023117738903.

Li, J., Yin, Y., Fortunato, S. and Wang, D. (2019), 'A dataset of publication records for Nobel laureates', *Scientific Data* 6(1), 1–10.

Merton, R. K. (1961), 'Singletons and multiples in scientific discovery: A chapter in the sociology of science', *Proceedings of the American Philosophical Society* 105(5), 470–486.

Mukherjee, S., Romero, D. M., Jones, B. and Uzzi, B. (2017), 'The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot', *Science Advances* 3(4), e1601315.

Pammolli, F., Magazzini, L. and Riccaboni, M. (2011), 'The productivity crisis in pharmaceutical R&D', *Nature Reviews Drug Discovery* 10(6), 428–438.

Petersen, A. M., Pan, R. K., Pammolli, F. and Fortunato, S. (2019), 'Methods to account for citation infla- tion in research evaluation', *Research Policy* 48(7), 1855–1865.

Singh, V. K., Singh, p. , Karmakar, M., Leta, J. and Mayr, P. (2021), 'The journal coverage of web of science,

scopus and dimensions: A comparative analysis', *Scientometrics* 126(6), 5113–5142.
Uzzi, B., Mukherjee, S., Stringer, M. and Jones, B. (2013), 'Atypical combinations and scientific impact', *Science* 342(6157), 468–472.
Waltman, L. (2016), 'A review of the literature on citation impact indicators', *Journal of informetrics* 10(2), 365–391.
Waltman, L. and van Eck, N. J. (2019), Field normalization of scientometric indicators, *in* 'Springer handbook of science and technology indicators', Springer, pp. 281–300.
Wang, D., Song, C. and Barabási, A.-L. (2013), 'Quantifying long-term scientific impact', *Science* 342(6154), 127–132.
Wang, J., Veugelers, R. and Stephan, P. (2017), 'Bias against novelty in science: A cautionary tale for users of bibliometric indicators', *Research Policy* 46(8), 1416–1436.
Wu, L., Wang, D. and Evans, J. A. (2019), 'Large teams develop and small teams disrupt science and technology', *Nature* 566(7744), 378–382.
Wuchty, S., Jones, B. F. and Uzzi, B. (2007), 'The increasing dominance of teams in production of knowledge', *Science* 316(5827), 1036–1039.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors did an excellent job with the revision of the documents. Before the manuscript can be published, I have four points that should be considered in another revision:

Main text, lines 60-61: Although some validation studies have been done in the past, the CD index cannot be seen as an established bibliometric indicator such as PP(top 10%). It is a relatively new indicator (proposed a few years ago) and further research is necessary to investigate its validity. For example, the indicator is based on an "ideal world" of citation decisions: only work is cited by authors that has (substantially) influenced the research reported in the focal paper and its citing papers. However, this ideal world does not exist. Bibliometric research has shown in many studies that many motives for citing exist such as to persuade readers of claims in a manuscript by citing reputable authors. An overview of this research can be found in Tahamtan & Bornmann (2018, 10.1016/j.joi.2018.01.002) and Tahamtan & Bornmann (2019, 10.1007/s11192-019-03243-4). The CD index may work well with patent data (an area where citations are more formalized than in publishing papers), but with bibliometric data some uncertainties remain. These uncertainties in the use of the CD index should be explained in the discussion part of the manuscript (as a limitation).

Main text, Figure 2: When thinking about the empirical results, another important factor came to my mind that should be controlled in the analyses. With respect to cited references, one must differentiate between linked and unlinked cited references in citation indexes (such as Web of Science). Linked cited references are connected to source documents in the database, and unlinked cited references are not. Unlinked cited references are mostly books, book chapters, and papers published in journals that are not covered in the database. There are differences between the disciplines in the share of unlinked cited references (e.g., the share is higher in the social sciences than in the life sciences) and between times. In general, one can expect over time that more cited references are linked. Thus, the decline in disruptiveness over time might be an effect of the increasing linking of cited references. Over time, more cited references appear as linked cited reference and thus as citation in the database. This linking effect should be observable over time (in all disciplines), and especially effective in certain disciplines such as social sciences. Thus, the found difference by the authors between disciplines and times may be (partly) driven by different shares of unlinked cited references.

Main text: In extended data Figure 9, results for some variants of the CD index are shown. Note that Bornmann et al. (2020, 10.1162/qss_a_00068) proposed to use a variant of the CD index where l>1 (at best l=5). Thus, the authors should switch to this variant in the figure.

Responses to R1C6: Thank you for this careful analysis and digging into the data. First, it seems that the fluctuations in the CD index are not data issues (as investigated by the authors). However, the authors should consider that the coverage of papers in the social sciences is generally bad (especially for earlier years). Only journal papers from core journals are included in the database (and books, book chapters, and other material not). Second, it is a problem of the statistical analysis, if the results cannot be qualitatively interpreted. Thus, it should be possible to dig into the data and to identify scientific activities that lead to the CD index fluctuations. If the connection between micro- and macro processes cannot be made, the results might be meaningless.

Referee #2 (Remarks to the Author):

Re-Review of “the decline of disruptive science and technology” by Park, Leahey and Funk for Nature.

In “the decline of disruptive science and technology” the authors use multiple measures, including an interesting one that they recently invented (and has received widespread attention), to trace a declining trend in science and technology over the latter half of the 20th Century. Their revision continues to analyze this in the context of citation and word-use patterns across multiple datasets, attempting to control for confounding explanations, and tries to identify explanations, painting an increasingly robust and coherent portrait of science and technology institutions that have slowed as they have grown, aged and narrowed. Their addition of multiple metrics has sharpened (and clarified) this revision against the prior one, and I believe that the removal of the crowd-sourced survey avoids distraction from the main findings.

There are even more things to like about the manuscript than before, including the careful, redundant measurement across fields (and even domains, e.g., JSTOR vs. the Web of Science us. USPTO). I believe that they have made a good faith and largely successful effort concerning measurement and effect identification. I still have questions about what this finding means—but there is no question that its going down.

On the interpretation, framing, and recommendation at the end, I was persuaded by their memo, but not their largely unrevised paper. Their “policy” recommendation that researchers personally write less, and “think” and “read more” seems both underspecified and mildly insulting. Are scientists not contributing to their new CD5 index because they are stupid? Why are they writing a lot of papers? Could there be incentives to write papers and build on work that everyone else also references? What might change these incentives to change the flow of attention and broaden/deepen science? Are there funding or institutional shifts that could help change this? This direction of reasonable policy-relevant thinking is not currently present in the discussion/conclusion.

Relatedly, the authors give lip service to the conservation of attention (the fixed number of disruptive papers has remained stable) in the memo (this is very interesting), but do not mention it (the conservation) in the discussion/conclusion. Instead, they mention that it is stable and optimistically suggest that we can change things by reflective reading and thinking. I think it would be more responsible, and true to the data analysis present, for them to suggest that this is a plausible explanation—that there could be a fixed carrying capacity in science, which might mean that it would be hard to change, at least without more dramatic institutional shifts and incentives. Consider Thomas Picketty’s famous book about Capital and Inequality. His book and evidence suggested that greater inequality was a natural, possibly inevitable, result of mature capital. But when he went on the road show for the book, he was pushed to put forward all kinds of government recommendations for fixing the trend—recommendations that were untested and without evidence—which ironically reduced the credibility of the main finding. I'm really overstating my point; I just think that a more nuanced and even-handed discussion at the end would give their paper more lasting-power.

Good job overall!

Referee #3 (Remarks to the Author):

As this is a second review. I am not going to reiterate points A-H which I covered in my first review.

The authors have responded substantively and in depth to the review team's observations. I am impressed with the work that went into this revision.

Although in my first review, I did not focus on the word analysis as much as the other reviewers, I too found it weak and confusing in the first version and was very glad to see that section has been simplified and in my view improved greatly.

Personally, I would prefer the Bornmann version of the indicator. But at this point, with so many supporting analyses in place, this is a matter of taste, and the manuscript presents a compelling case in its current form.

The authors ask me about their new version of the randomness check. I appreciate the changes made. The analysis and rationale for it now makes sense to me. So I agree with including it.

Author Rebuttals to First Revision:

Revisions Made in Response to Comments ("C") from Reviewer ("R") 1

R1C1: *The authors did an excellent job with the revision of the documents. Before the manuscript can be published, I have four points that should be considered in another revision:*

Thank you for your kind words on our revised manuscript. We greatly appreciate your helpful feedback throughout the review process.

R1C2: *Main text, lines 60-61: Although some validation studies have been done in the past, the CD index cannot be seen as an established bibliometric indicator such as PP(top 10%). It is a relatively new indicator (proposed a few years ago) and further research is necessary to investigate its validity. For example, the indicator is based on an "ideal world" of citation decisions: only work is cited by authors that has (substantially) influenced the research reported in the focal paper and its citing papers. However, this ideal world does not exist. Bibliometric research has shown in many studies that many motives for citing exist such as to persuade readers of claims in a manuscript by citing reputable authors. An overview of this research can be found in Tahamtan & Bornmann (2018, 10.1016/j.joi.2018.01.002) and Tahamtan & Bornmann (2019, 10.1007/s11192-019-03243-4). The CD index may work well with patent data (an area where citations are more formalized than in publishing papers), but with bibliometric data some uncertainties remain. These uncertainties in the use of the CD index should be explained in the discussion part of the manuscript (as a limitation).*

Thank you for this suggestion. Based on your feedback, we added the following paragraph to the discussion section.

> *Our study is not without limitations. Notably, while research to date supports the validity of the CD index[12,29], it is a relatively new indicator of innovative activity and will benefit from future work on its behavior and properties, especially across data sources and contexts. Studies that systematically examine the effect of different citation practices[52,53], which vary across fields, would be particularly informative.*

R1C3: *Main text, Figure 2: When thinking about the empirical results, another important factor came to my mind that should be controlled in the analyses. With respect to cited references, one must differentiate between linked and unlinked cited references in citation indexes (such as Web of Science). Linked cited references are connected to source documents in the database, and unlinked cited references are not. Unlinked cited references are mostly books, book chapters, and papers published in journals that are not covered in the database. There are differences between the disciplines in the share of unlinked cited references (e.g., the share is higher in the social sciences than in the life sciences) and between times. In general, one can expect over time that more cited references are linked. Thus, the decline in disruptiveness over time might be an effect of the increasing linking of cited references. Over time, more cited references appear as linked cited reference and thus as citation in the database. This linking effect should be observable over time (in all disciplines), and especially effective in certain disciplines such as social sciences. Thus, the found difference by the authors between disciplines and times may be (partly) driven by different shares of unlinked cited references.*

Thank you for this suggestion. Following your input, we added a control for the number of unlinked references to our regression models that adjust the CD index for changing publication, citation, or authorship practices (see discussion in the Methods section, specifically the subsection “Verification using regression adjustment,” along with the results in Supplementary Table 1 and Extended Data Figure 8b and Figure 8e). We continue to find similar results after including this control, which offers reassurance that our findings are not driven by changes in the relative prevalence of linked/unlinked references over time.

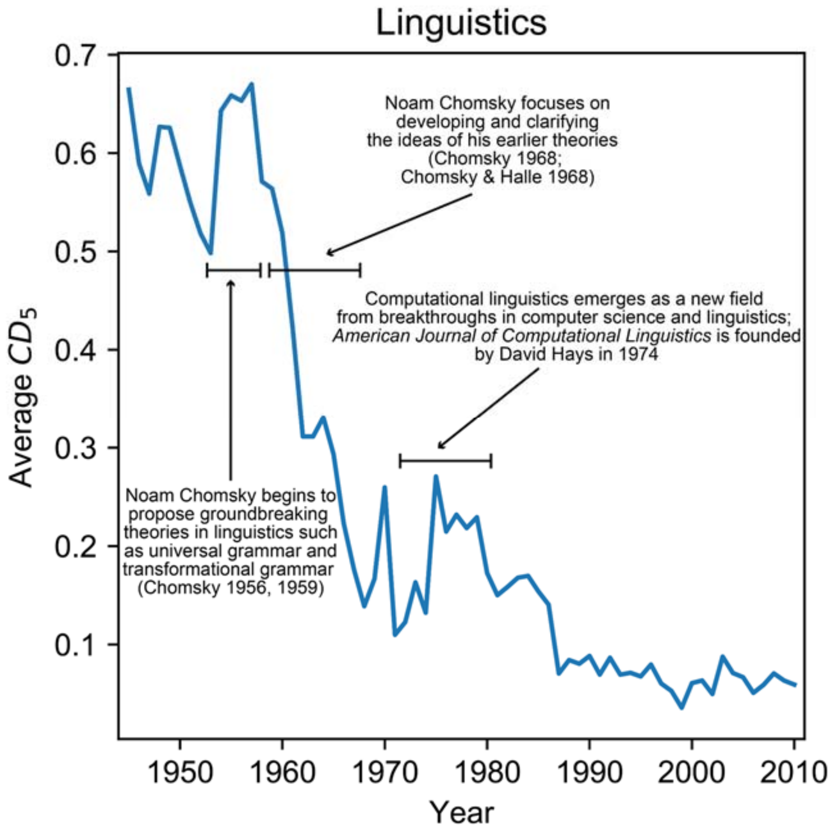
R1C4: *Main text: In extended data Figure 9, results for some variants of the CD index are shown. Note that Bornmann et al. (2020, 10.1162/qss_a_00068) proposed to use a variant of the CD index where l>1 (at best l=5). Thus, the authors should switch to this variant in the figure.*

Thank you for this suggestion. Following your input, we changed the figure (now Extended Data Figure 7) to the l=5 variation of the measure proposed by Bornmann et al. (2020). We find similar results to those reported in the previous version of the manuscript.

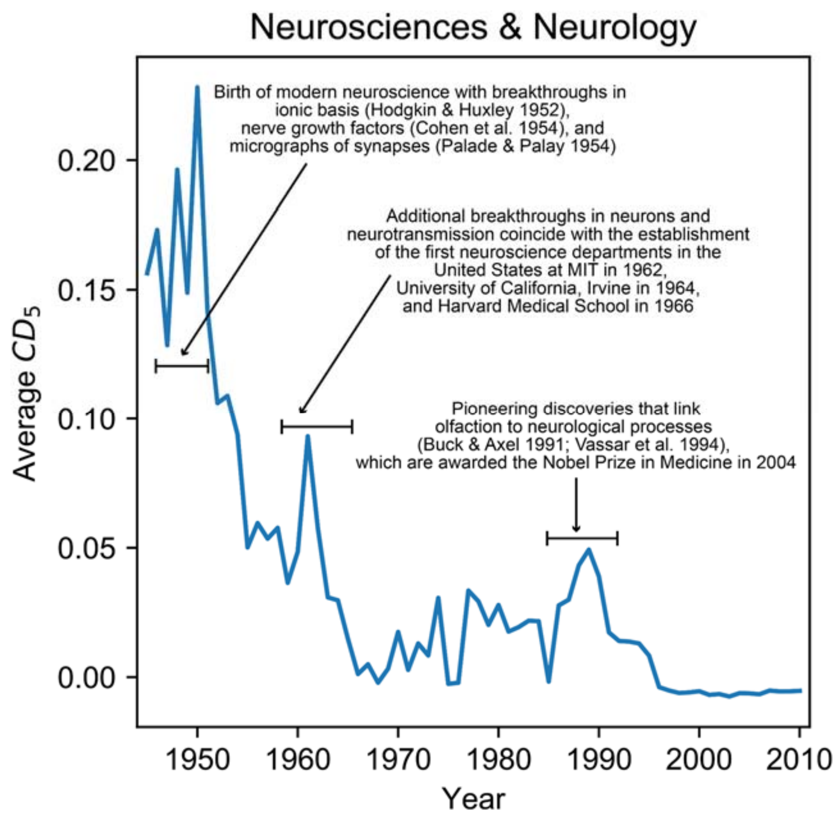
R1C5: *Responses to R1C6: Thank you for this careful analysis and digging into the data. First, it seems that the fluctuations in the CD index are not data issues (as investigated by the authors). However, the authors should consider that the coverage of papers in the social sciences is generally bad (especially for earlier years). Only journal papers from core journals are included in the database (and books, book chapters, and other material not). Second, it is a problem of the statistical analysis, if the results cannot be qualitatively interpreted. Thus, it should be possible to dig into the data and to identify scientific activities that lead to the CD index fluctuations. If the connection between micro- and macro processes cannot be made, the results might be meaningless.*

Thank you for your positive comments on our revisions from the prior round. Regarding your first concern about the coverage of social sciences, while we agree that this is a limitation of the WoS database, we believe there are two considerations that help allay this concern for our study. First, as mentioned in a previous response (**R1C3**), we find consistent results when running regressions that control for the number of linked/unlinked references over time. Second, we observe consistent declines in disruptiveness over time across multiple different databases, including those that have much more robust coverage of the social sciences than WoS. To this point, in Extended Data Figure 6 we continue to find a consistent decline in disruptiveness among papers archived in the JSTOR database, which is known for its wide coverage of social science journals. Moreover, while JSTOR does not reliably index non-journal publications, coverage of non-journal sources is much better in Microsoft Academic Graph, and there too, we see a consistent decline in disruptiveness.

Following your comment, we also dug more deeply into the various fluctuations in the CD index over time, with the goal of verifying that the aggregate patterns we observe in the measure correspond with concrete changes in scientific activities. To do so, we conducted case studies of several different substantive fields, and found that significant fluctuations often correspond to identifiable historical events. Here, we highlight two fields, corresponding to the WoS subject areas of “Linguistics” and “Neuroscience and Neurology.” We focused on these two fields because they: (1) represent two diverse areas of science (social sciences and life sciences), (2) had notable fluctuations in their average CD index values over time, and (3) have fairly well documented records of major historical events and scientific activities. The results of our case studies are summarized in the figures and descriptions below.



In linguistics, we find that Noam Chomsky's groundbreaking work, which proposed new theories on universal and transformational grammars, contributed to increases in disruption between 1950-1960. Thereafter, the field, including Chomsky, focused on developing and clarifying these seminal theories. Finally, between 1970-1980, breakthroughs in computer science and linguistics laid the groundwork for the emergence of computational linguistics as a new area of study, which coincided with the founding of the *American Journal of Computational Linguistics* in 1974 (Grishman 1986; Barsky 1997).



Key disruptive breakthroughs around 1950 laid the foundations for the emergence of modern neuroscience. Additional groundbreaking discoveries occurred in the 1960s, which coincided with the establishment of the first neuroscience departments at U.S. institutions such as MIT, UC Irvine, and Harvard. Finally, disruptive discoveries linking olfaction and neurological processes occurred in the 1990s, which were subsequently awarded the Nobel Prize in Medicine in 2004 (Gross 1999; Cowan, Harter & Kandel 2000).

Revisions Made in Response to Comments ("C") from Reviewer ("R") 2

R2C1: *Re-Review of "the decline of disruptive science and technology" by Park, Leahey and Funk for Nature.*

In "the decline of disruptive science and technology" the authors use multiple measures, including an interesting one that they recently invented (and has received widespread attention), to trace a declining trend in science and technology over the latter half of the 20th Century. Their revision continues to analyze this in the context of citation and word-use patterns across multiple datasets, attempting to control for confounding explanations, and tries to identify explanations, painting an increasingly robust and coherent portrait of science and technology institutions that have slowed as they have grown, aged and narrowed. Their addition of multiple metrics has sharpened (and clarified) this revision against the prior one, and I believe that the removal of the crowd-sourced survey avoids distraction from the main findings.

There are even more things to like about the manuscript than before, including the careful, redundant measurement across fields (and even domains, e.g., JSTOR vs. the Web of Science us. USPTO). I believe that they have made a good faith and largely successful effort concerning measurement and effect identification. I still have questions about what this finding means—but there is no question that its going down.

Thank you for your positive comments on our work. We greatly appreciate your helpful comments and suggestions throughout the review process.

R2C2: *On the interpretation, framing, and recommendation at the end, I was persuaded by their memo, but not their largely unrevised paper. Their "policy" recommendation that researchers personally write less, and "think" and "read more" seems both underspecified and mildly insulting. Are scientists not contributing to their new CD5 index because they are stupid? Why are they writing a lot of papers? Could there be incentives to write papers and build on work that everyone else also references? What might change these incentives to change the flow of attention and broaden/deepen science? Are there funding or institutional shifts that could help change this? This direction of reasonable policy-relevant thinking is not currently present in the discussion/conclusion.*

Thank you for drawing this concern to our attention. In light of your feedback, we have rewritten the discussion section. Specifically, we dropped our previous recommendations on reading and writing strategies for researchers. Instead, we now focus our recommendations on institutional changes that might help better support researchers and create conditions that are more conducive to the production of disruptive work.

R2C3: *Relatedly, the authors give lip service to the conservation of attention (the fixed number of disruptive papers has remained stable) in the memo (this is very interesting), but do not mention it (the conservation) in the discussion/conclusion. Instead, they mention that it is stable and optimistically suggest that we can change things by reflective reading and thinking. I think it would be more responsible, and true to the data analysis present, for them to suggest that this is a plausible explanation—that there could be a fixed carrying capacity in science, which might mean that it would be hard to change, at least without more dramatic institutional shifts and incentives. Consider Thomas Picketty's famous book about Capital and Inequality. His book and evidence suggested that greater* *inequality was a natural, possibly inevitable, result of mature capital. But when he went on the road show for the book, he was pushed to put forward all kinds of government recommendations for fixing the trend—recommendations that were untested and without evidence—which ironically reduced the credibility of the main finding. I'm really overstating my point; I just think that a more nuanced and even-handed discussion at the end would give their paper more lasting-power.*

Thank you for this suggestion. As noted in our previous response, we have rewritten the Discussion section in light of your feedback. In addition to offering a new set of more institutionally-oriented policy implications, we also address the implications of our findings on the conservation of highly disruptive work more directly.

R2C4: *Good job overall!*

Thank you for your kind words on our revision, and for your helpful feedback throughout the review process. We know our paper is much stronger as a result of your input.

Revisions Made in Response to Comments (“C”) from Reviewer (“R”) 3

R3C1: *As this is a second review. I am not going to reiterate points A-H which I covered in my first review.*

The authors have responded substantively and in depth to the review team's observations. I am impressed with the work that went into this revision.

Thank you for your positive comments on our revision.

R3C2: *Although in my first review, I did not focus on the word analysis as much as the other reviewers, I too found it weak and confusing in the first version and was very glad to see that section has been simplified and in my view improved greatly.*

Thank you. We are glad to hear our revisions have improved the clarity of the manuscript.

R3C3: *Personally, I would prefer the Bornmann version of the indicator. But at this point, with so many supporting analyses in place, this is a matter of taste, and the manuscript presents a compelling case in its current form.*

Thank you for your positive comments.

R3C4: *The authors ask me about their new version of the randomness check. I appreciate the changes made. The analysis and rationale for it now makes sense to me. So I agree with including it.*

Thank you very much for your input. We will keep the simulation analysis in the manuscript.

In closing, thank you once again for your helpful feedback throughout the review process. We know our paper is much stronger as a result of your input.

References

Barsky, R. F. (1998). *Noam Chomsky: A life of dissent*. MIT Press.

Buck, L., & Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, *65*(1), 175-187.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, *2*(3), 113-124.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*(2), 137-167.

Chomsky, N. (1968). *Remarks on nominalization*. Linguistics Club, Indiana University.

Chomsky, N., & Halle, M. (1968). The sound pattern of English, New York: Harper & Row.

Cohen, S., Levi-Montalcini, R., & Hamburger, V. (1954). A nerve growth-stimulating factor isolated from sarcom as 37 and 180. *Proceedings of the National Academy of Sciences*, *40*(10), 1014-1018.

Cowan, W. M., Harter, D. H., & Kandel, E. R. (2000). The emergence of modern neuroscience: some implications for neurology and psychiatry. *Annual Review of Neuroscience*, *23*, 343.

Grishman, R. (1986). *Computational linguistics: an introduction*. Cambridge University Press.

Gross, C. G. (1999). *Brain, vision, memory: Tales in the history of neuroscience*. MIT Press.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, *117*(4), 500.

Palay, S. L., & Palade, G. E. (1955). The fine structure of neurons. *The Journal of Biophysical and Biochemical Cytology*, *1*(1), 69.

Vassar, R., Chao, S. K., Sitcheran, R., Nun, J. M., Vosshall, L. B., & Axel, R. (1994). Topographic organization of sensory projections to the olfactory bulb. *Cell*, *79*(6), 981-991.

Editorial Note:

These responses were sent back to the three referees and all recommended publication of the revised paper.