Digital Scholarship in the Humanities, 2024, **39**, 723–735
https://doi.org/10.1093/llc/fqae004
Advance access publication 8 February 2024
Full Paper

OXFORD

Whose Anthropocene?: a data-driven look at the prospects for collaboration between natural science, social science, and the humanities

Carlos Santana [1,*], **Kathryn Petrozzo**[2], **T.J. Perkins** [2]

[1]Department of Philosophy, University of Pennsylvania, Philadelphia, Pennsylvania, United States
[2]Department of Philosophy, University of Utah, Salt Lake City, Utah, United States

*Corresponding author. Department of Philosophy, University of Pennsylvania, Philadelphia, Pennsylvania, United States.
E-mail: csantana@sas.upenn.edu

Abstract

Although the idea of the Anthropocene originated in the earth sciences, there have been increasing calls for questions about the Anthropocene to be addressed by pan-disciplinary groups of researchers from across the natural sciences, social sciences, and humanities. We use data analysis techniques from corpus linguistics to examine academic texts about the Anthropocene from these disciplinary families. We read the data to suggest that barriers to a broadly interdisciplinary study of the Anthropocene are high, but we are also able to identify some areas of common ground that could serve as interdisciplinary bridges.

Keywords: Anthropocene; interdisciplinarity; corpus analysis; humans and nature; global change.

1. Introduction

Although it originated in the earth sciences (Crutzen and Stoermer 2000), the concept *Anthropocene* has entered widespread use across academic disciplines and even in public discourse. This widespread adoption of the term has occurred even though the Anthropocene has not yet become a recognized unit of the geologic time scale (GTS). Popular use of the concept even long predated the official recommendation of the Anthropocene Working Group that it be considered for inclusion in the GTS (Zalasiewicz *et al*. 2019). Given this apparent gap between what the term "Anthropocene" means in the geosciences, and what it means elsewhere, there have been numerous calls for interdisciplinary collaboration in defining, designating, and researching the Anthropocene, involving disciplines such as history (Chakrabarty 2009), the environmental humanities (Castree 2014), anthropology (Lock 2018), and the humanities and social sciences in general (Thomas 2014). There have even been calls in leading scientific journals like *Nature* to revise the standard procedures for ratifying units of the GTS, and ensure that social scientists, ecologists, and philosophers have a say in the eventual decision about the Anthropocene (Ellis *et al*. 2016).

What are the prospects for this sort of unified research on the Anthropocene? Interdisciplinary collaboration is always easier said than done. For example, a recent collaboration between a historian and natural scientists characterizes itself as a "multi-disciplinary" approach to the Anthropocene rather than an interdisciplinary one, because while the various disciplines may be interested in the same issue in a broad sense, we have not yet been able to synthesize and harmonize work from across the disciplines to answer shared questions (Thomas, Williams, and Zalasiewicz 2020).

The research presented here is a data-driven approach to understanding the commonalities and barriers between different disciplines interested in the Anthropocene. We aim to expose the barriers to cross-discipline communication about the Anthropocene, as well as to identify potential common ground between geoscience, social science, and humanities researchers studying the Anthropocene. To do so, we follow recent work in the empirical philosophy of science in using computational methods on corpora of academic text (e.g. Pence and Ramsey 2018; Hicks 2021; Mizrahi and Dickinson 2022). We generated three corpora for analysis, each consisting of journal articles about the Anthropocene from each family of disciplines. This

allows us to compare and contrast what each discipline means when they use the term "Anthropocene," as well as explore what matters to each discipline in research about the Anthropocene. Drawing on our results, we adopt a measured pessimism about cross-academy collaboration on Anthropocene research, given that our findings indicate that the shared terminology of "Anthropocene" hides divergent concepts unique to each disciplinary family.

2. Corpus construction

We assembled three corpora, one each from the geosciences, social sciences, and humanities. Each corpus sampled articles from a set of journals in the appropriate disciplines (see Appendix 1 for a full list of journals). Journals were chosen from among those highly ranked in their field according to SCImago Journal Rank (SJR), with preference given to journals with high numbers of articles containing the token "Anthropocene." For the social science corpus, care was given to sample journals from across the major social science fields, including economics, psychology, anthropology, sociology, and geography.[1] A final consideration was to have each corpus roughly equal in size, and so some journals with a small number of articles sampled were added to obtain this balance.

Each corpus contains the full text of every article available online from the selected journals containing the token "Anthropocene," as of August 2019 (Geosciences and Humanities corpora) or October 2019 (Social Science corpus). Articles not containing the token "Anthropocene" were not included in the corpora. Each article was manually cleaned to remove repetitive information in headers and footers, including journal titles, author names, and article titles. Each of the three resulting corpora contains approximately 1.5 million words.[2]

3. Data analysis

To assess both common ground and differences among the three corpora, we conducted three types of analyses: keywords, collocates, and topic modeling.

All three analyses were case-insensitive (e.g. we treated "Climate" and "climate" as equivalent tokens).

3.1 Keywords

Keywords are words which are distinctive of a corpus in contrast to another corpus. We identified keywords using Kilgarriff's Simple Maths Procedure (SMP) with $k = 100$ (Kilgarriff 2009; see Box 1). The higher the SMP for a keyword, the more that word is distinctive of a focal corpus in relation to a contrast corpus. For instance, with our Geosciences corpus as focal, and our Humanities corpus as contrast, the word *sediment* scores an extremely high SMP of 16.83, suggesting that discussion of sediment is one of the strongest features distinguishing the Geosciences corpus from the Humanities corpus. To ensure that an SMP measure is not skewed by a word being highly frequent in a small subset of documents, we also measured *dispersion*, which is roughly how evenly spread out a word is among documents in a corpus (Kilgarriff 1997). To measure dispersion, we used the coefficient of variation (CV; see Box 1). The lower the CV, the more evenly dispersed the word throughout the corpus, and thus the more likely the corresponding keyword measure is not a result of a biased sample. There is no standard cutoff when using dispersion to assess significance, but for corpora of this size, a CV in the single digits suggests that the keyword measure is almost certainly reliable. For instance, the dispersion of the term *sediment* in the Geosciences corpus is 1.88, meaning that it is used nearly evenly across documents in the corpus and thus the SMP measure for *sediment* is highly reliable. Keywords capture any lexical differences across corpora, and so some of the keywords we find will represent general stylistic differences between disciplines, not differences in how the Anthropocene in particular is discussed.[3]

Box 1. Keyword measures

Kilgarriff's SMP:

$$\frac{\text{frequency (per 10k words) in focal corpus} + \text{k}}{\text{frequency (per 10k words) in contrast corpus} + \text{k}}$$

CV:

$$\frac{\textit{standard deviation (relative frequency per document)}}{\textit{mean (relative frequency per document)}}$$

In our results, we certainly do see keywords that arise from general stylistic differences. For example, "et" and "al." are keywords in the geoscience corpus when contrasted with the humanities, which is a predictable result of the fact that co-authorship is more common in the earth sciences than the humanities. Seeing these broad disciplinary differences reflected in the keywords is encouraging since it shows that the measure we are using is picking up on real effects. However, as will become clear when we get to the analysis, our keyword results are also capturing the variegated structure of Anthropocene-specific discourse. A wide range of the keywords are clearly driven by the emphasis on the Anthropocene. For instance, several keywords refer to authors who have written on the topic (e.g. "Haraway," "Latour"), and others speak to the parts of disciplines that relate more closely to the Anthropocene (e.g. "Quaternary" for geoscience and "sustainability" for social science). Moreover, our interpretation of the keyword data is done in the context of the collocate data, which is specific to the immediate linguistic context of the term "Anthropocene." So, we are confident that the keywords, taken as a set and appropriately contextualized using the other data, speak to how disciplines think about the Anthropocene, and not just how writing differs between fields in general.

Complete frequency, keyword, and dispersion data are available in Supplementary data files. In Tables 1–4, we present the top ninety-nine keywords with dispersion less than ten for the Geosciences corpus contrasted with the other two corpora.

3.2 Collocates

Collocates are words appearing within a specified range of a word of interest. For our analysis, we compared the collocates of the word "Anthropocene" in each corpus, using the *Lancsbox X* software package (Brezina and Platt 2023) to generate positional graphs of the most frequent collocates within a span of five words to the left and right of "Anthropocene." In a positional graph, the length of the edge (line segment) from a collocate to the node ("Anthropocene") represents the strength of association between the two, measured using logDice (Rychlý 2008). That is, collocates with shorter edges are more likely to appear near "Anthropocene" than elsewhere in the corpus. Position to the left or right of the node represents the mean distance from the node in the text, for example, a collocate positioned far to the left on the figure appears on average 3–4 words prior to "Anthropocene" in the corpus. And the size of a collocate's node (circle) represents its raw collocation frequency, with larger nodes representing items collocated with Anthropocene more frequently. Color represents frequency in the corpus in general, with hotter (redder) colors representing more frequent words.

Table 1. Geoscience keywords in contrast with Humanities

Word	SMP ($k = 100$)
Al	18.69
Et	18.23
sediment	16.83
holocene	13.21
sediments	10.91
m	10.46
bp	10.17
basin	9.82
fig	9.39
quaternary	8.00
ka	7.67
c	7.64
co2	7.08
data	6.95
records	6.94
erosion	6.84
pollen	6.73
lake	6.71
b	6.57
sci	6.44
deposition	6.33
record	6.17
samples	6.07
temperature	6.04
concentrations	6.03
during	6.01
lower	6.01
j	5.91
carbon	5.88
vegetation	5.78
table	5.69
deposits	5.68
catchment	5.61
increase	5.59
changes	5.55
river	5.41
cm	5.38
variability	5.30
peat	5.28
sedimentary	5.18
lakes	5.07
rates	5.01
shelf	4.99
precipitation	4.93
china	4.69
dating	4.67
atmospheric	4.65
upper	4.60
northern	4.59
period	4.54
region	4.52
fluvial	4.51
high	4.50
model	4.46
ad	4.46
e.g.	4.45
cal	4.39
archaeological	4.39
p.	4.37

(continued)

Table 1. (continued)

Word	SMP ($k = 100$)
gravel	4.37
isotope	4.35
area	4.34
maximum	4.30
charcoal	4.30
v.	4.29
glacial	4.27
radiocarbon	4.24
low	4.23
pp	4.23
cover	4.23
monsoon	4.21
±	4.21
continental	4.17
tropical	4.15
increased	4.14
total	4.11
depth	4.10
gulf	4.09
values	4.09
using	4.07
drainage	4.05
km	4.02
rate	4.02
res	4.00
r	3.97
mississippi	3.95
n	3.94
carbonate	3.92
marine	3.90
climatic	3.90
soil	3.86
ages	3.86
proxy	3.84
ce	3.83
sedimentation	3.83
estimates	3.81
core	3.81
channel	3.81
drought	3.79

Figures 1–3 present collocate information for the 100 most frequent collocates for each of the three corpora. In each figure, we have removed function words, numerals, and words like "doi" and "press" that appear only due to how common they are in reference lists (author names are retained). To enhance legibility we have made minimal adjustments to node position, retaining edge length and left-right position as much as possible, when two collocates overlapped and were unreadable.

3.3 Topics

Topics are sets of words which cluster together in a subset of documents. We combined all three disciplinary corpora into a single corpus, and then determined topics using Latent Dirichlet Allocation (LDA; Blei, Ng, and Jordan 2003) using the Mallet machine learning software

Table 2. Humanities keywords in contrast with Geoscience

Word	SMP ($k = 100$)
her	11.61
life	10.27
his	10.25
she	10.01
humanities	8.58
what	8.16
he	7.19
you	7.12
my	6.81
sense	6.50
politics	6.40
things	6.36
own	6.21
who	5.76
space	5.42
us	5.28
ways	5.18
i	5.15
thinking	5.09
nonhuman	5.00
think	4.86
political	4.84
literary	4.78
narrative	4.73
like	4.64
social	4.60
violence	4.58
experience	4.57
animals	4.52
world	4.49
something	4.49
art	4.47
edited	4.35
how	4.34
ethics	4.32
kind	4.29
story	4.27
itself	4.20
poetry	4.17
way	4.16
care	4.16
stories	4.14
bodies	4.10
them	4.02
death	4.02
relations	4.00
me	3.99
living	3.94
beings	3.93
ecocriticism	3.92
place	3.89
war	3.85
man	3.84
worlds	3.82
writing	3.82
culture	3.80
haraway	3.77
animal	3.77
media	3.74

(continued)

Table 2. (continued)

Word	SMP ($k = 100$)
lives	3.73
accessed	3.73
poem	3.71
ibid	3.70
ethical	3.63
making	3.62
out	3.57
your	3.57
might	3.56
fiction	3.55
language	3.50
make	3.49
creative	3.48
york	3.44
essay	3.44
thought	3.43
him	3.43
work	3.42
nature	3.40
body	3.39
film	3.38
agency	3.37
geographies	3.35
people	3.35
forms	3.34
discourse	3.33
narratives	3.33
latour	3.33
it	3.32
reading	3.31
humanity	3.30
multispecies	3.29
call	3.29
toward	3.29
literature	3.28
just	3.25
feminist	3.25
print	3.25
novel	3.24
attention	3.19

package (McCallum 2002). We generated twenty topics,[4] ignoring common but uninformative words like *the*, *or*, and *so*. LDA probabilistically assigns topics, but it does not interpret them, so each author independently labeled each topic, then we determined consensus labels based on our individual labels. Topic interpretation can be fraught, and at its worst has even been compared to the unscientific reading of tea leaves (Chang *et al*., 2009), but in the case of our Anthropocene corpora, topics were unusually clear. For instance, we labeled Topic 10, "Species Conservation," given that it included, among other words: *species*, *biology*, *conservation*, *invasive*, *biological*, *biodiversity*, *extinction*, and *invasion*. Similarly, the topic we labeled "Literature" includes in its top tokens *literature*, *literary*, *poetry*, *ecocriticism*, *poem*, *reading*, *print*, *fiction*, *story*, and *narrative*, while the

Table 3. Geoscience keywords in contrast with Social Science

Word	SMP (k = 100)
sediment	15.11
holocene	12.20
sediments	10.52
bp	9.29
lake	7.90
quaternary	7.66
ka	7.57
records	7.37
pollen	6.88
ice	6.76
record	6.42
temperature	6.04
basin	6.01
samples	5.99
lakes	5.79
concentrations	5.71
river	5.70
organic	5.48
cm	5.42
sea	5.36
ad	5.35
shelf	5.30
deposits	5.30
ocean	5.28
catchment	5.27
deposition	5.25
co2	5.12
sedimentary	5.09
atmospheric	5.04
m	5.01
figure	4.89
al	4.89
vegetation	4.88
peat	4.82
cores	4.79
et	4.70
dating	4.63
drought	4.58
fluvial	4.51
erosion	4.49
gravel	4.35
precipitation	4.32
during	4.31
cal	4.28
glacial	4.24
late	4.20
northern	4.17
monsoon	4.13
continental	4.09
j	4.07
isotope	4.06
radiocarbon	4.05
gulf	4.02
mississippi	4.01
channel	4.00
carbonate	3.97
terrestrial	3.95
surface	3.94
upper	3.94

(continued)

Table 3. (continued)

Word	SMP (k = 100)
tropical	3.91
±	3.90
ages	3.87
stratigraphic	3.86
marine	3.81
pleistocene	3.80
yr	3.73
sedimentation	3.73
variability	3.72
climatic	3.72
c	3.69
depth	3.66
rivers	3.60
temperatures	3.56
ce	3.56
snow	3.55
210pb	3.52
lower	3.52
methane	3.50
dates	3.49
valley	3.48
fluxes	3.48
southern	3.46
birks	3.45
flux	3.43
proxy	3.39
isotopic	3.39
14c	3.38
flood	3.37
anthropogenic	3.37
forcing	3.33
fan	3.32
age	3.30
concentration	3.25
core	3.25
proxies	3.24
floodplain	3.21
mediterranean	3.20
dust	3.18
res	3.17

"Dating" topic's full top ten tokens are *core*, *peat*, *dating*, *age*, *ice*, *Anthropocene*, *sediments*, *radiocarbon*, *cores*, and *sediment*. A skeptic would have a hard time arguing that the labels we assigned to these topics are arbitrary or esoteric. The other seventeen topics were similarly transparent in their interpretation, and there was consensus among all authors on the labels for each topic. Complete topic data are available in the supplementary material.

LDA topic modeling assigns topic weights for each topic to each document in the corpus, with weights summing to one. For example, Hayes *et al.* (2017) receive a weight of 0.44 for the topic "Dating," 0.23 for "Ocean Geology," and 0.16 for "Contemporary Climate Change," with the remaining topics receiving insignificant weights. The title of the paper is "Helium and

Table 4. Social Science keywords in contrast with Geoscience

Word	SMP ($k = 100$)
social	10.97
anthropology	9.29
economics	9.05
political	6.94
politics	6.21
governance	5.90
energy	5.76
life	5.73
her	5.63
economic	5.53
what	5.48
she	5.44
care	5.34
anthropologists	5.16
public	5.14
who	5.01
how	5.01
anthropological	4.93
theory	4.66
capital	4.66
people	4.60
ways	4.52
institutional	4.44
health	4.29
his	4.23
relations	4.16
he	4.14
capitalism	4.11
institutions	4.11
econ	4.08
own	4.06
power	3.98
sociology	3.96
costs	3.93
economy	3.93
2018	3.90
ethnographic	3.88
policy	3.84
sense	3.79
practice	3.75
my	3.74
practices	3.73
market	3.70
consumption	3.70
accessed	3.65
everyday	3.61
ethical	3.54
legal	3.54
culture	3.53
work	3.53
collective	3.43
sustainability	3.43
justice	3.42
attention	3.40
ethnography	3.37
law	3.37
anthropologist	3.36
lives	3.34
income	3.34

(continued)

Table 4. (continued)

Word	SMP ($k = 100$)
action	3.31
things	3.30
knowledge	3.27
them	3.25
individuals	3.25
agency	3.24
futures	3.23
women	3.20
medical	3.19
rights	3.16
actors	3.16
think	3.15
i	3.12
race	3.11
way	3.10
&	3.10
us	3.08
media	3.08
violence	3.08
urban	3.08
indigenous	3.06
goods	3.04
scholars	2.97
ecological	2.94
world	2.92
cultural	2.91
themselves	2.89
make	2.88
resource	2.88
policies	2.87
forms	2.85
markets	2.82
commons	2.81
thinking	2.79
making	2.79
ethics	2.75
you	2.75
money	2.75
critique	2.75
others	2.75

thorium isotope constraints on African dust transport to the Bahamas over recent millennia," so this assignment of topic weights seems accurate. Complete topic weight assignments are available in the supplementary material. Here, since we cannot present weights for hundreds of individual documents, we will compare mean topic weights between corpora, which should allow us to identify areas where the disciplines overlap and diverge in topics. Table 5 presents the mean topic weight by corpus as a heatmap.

4. Interpreting the data

First, a word on our approach to interpretation. The corpus methods we employ are illuminating, but as with any data-driven method, the evidence is always an incomplete representation of the phenomena of interest. Additionally, we have had to make decisions about which methods to use (which keyword measure, what topic modeling algorithm, and so on). These decisions are about navigating tradeoffs, as there is generally no single best method for corpus analysis. It is for this reason that we are using three different, independent methods, and in our interpretations look for areas where multiple methods inform the lesson we draw.

Furthermore, the set of lessons we draw is not meant to be comprehensive or the last word. We hope that keen readers draw on our figures, tables, and Supplementary data to answer questions we did not even think to ask. And we hope that researchers who would quibble with our techniques apply their own favored methods to our corpus to push our understanding of Anthropocene scholarship forward.

Now that the preliminaries are out of the way, we can assert that our results paint a clear picture of disciplinary disconnect. To a certain extent, this is unsurprising. It is well understood that different disciplines differ in method, terminology, and conceptual framing devices. For instance, many researchers have noted that time scales differ in both degree and kind between the earth sciences and the social sciences (Chakrabarty 2009; Thomas, Williams, and Zalasiewicz 2020). And no one is surprised that, say, economists typically use different techniques and attend to different evidence than atmospheric scientists. Our data shed light on differences of these sorts, of course, but also reveal a less obvious but substantive divide: we interpret the data as showing that the term "Anthropocene" means something quite different in each disciplinary family. It is not merely that each discipline is researching the same object from different perspectives, but that there is not much of a shared object of research at all. It is not only that the disciplines do not share a vocabulary, method, or even a set of questions. Research on the Anthropocene in different disciplines does not even necessarily target the same general objects and issues. Consequently, the barriers to pan-disciplinary Anthropocene studies are substantial. In this section, we will look in detail at the linguistic data to better understand the barriers, and to highlight opportunities for overcoming them.

Interdisciplinarity is easier to achieve when methods differ but goals align, but that is not what we see in these corpora. In our assessment of the data, the different disciplines do not have the same goals, and so are not asking the same questions. Collocate data (Fig. 1) demonstrate that geoscience research on the Anthropocene is dominated by questions of (geo)chronology: When is the *start*, *onset*, or *beginning* of the Anthropocene? What type of *unit* should it receive

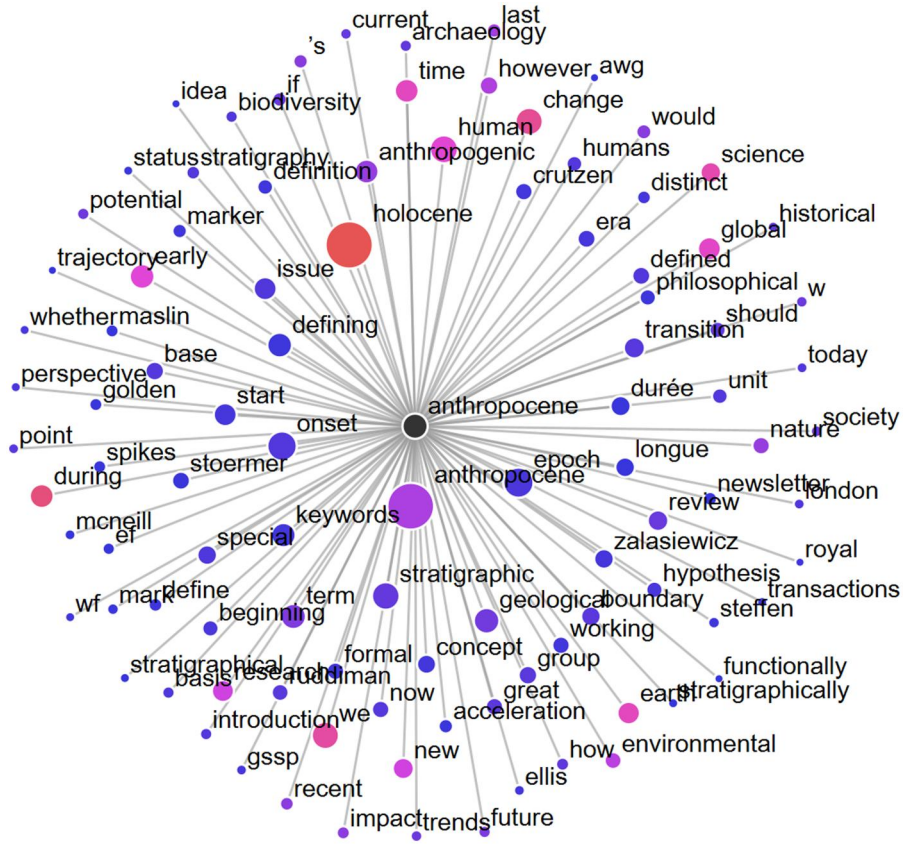

Figure 1. Geoscience collocates.

formal definition as, an *epoch*, *period*, or *era*? Humanities research on the Anthropocene does not seem to take up these questions, as evidenced by both keywords and collocates. Instead, humanities scholars are concerned with finding the right *term* and *concept*. Is the name "Anthropocene" appropriate, or is it better labeled *Capitalocene, Plantationocene*, or *Cthulucene*, as suggested by the frequently cited *Donna Haraway* (2015)? Responsibility for global change is not equally apportioned among all of humanity, so humanities scholars are critical of the label "Anthropocene," which lumps all of humanity together. Finally, topic modeling (Table 5) indicates that social science research on the Anthropocene is policy oriented, with economic policy, environmental justice, conservation, and indigenous policy being the strongest threads in the Social Science corpus. Topic data show that the Humanities corpus shares some of these emphases, to a lesser extent, but very little of the Geoscience corpus touches on the policy questions which dominate the social sciences. Summarized simply, earth scientists are asking "Is there an Anthropocene, and if so, when did it begin?" while the other disciplines are taking its existence for granted and then asking "How should we conceptualize it?" "What should we do about it?" and "How does it impact us?".

While these research questions are distinct, they potentially feed into each other. But from what we can see in the corpus, this potential overlap rarely becomes actual. For instance, the processes of interest to earth scientists (sedimentation, hydrology, and atmospheric change) and the processes of interest to social scientists (economic, demographic, and institutional change) are deeply entwined in the modern world, and so there is an opportunity to bridge these fields by researching the interaction between these human and earth-system processes. We could be asking, for instance, about how hydrology interacts with economic growth, or how demographic shifts affect atmospheric composition. While such interdisciplinary research exists, the corpora indicate that it is not a significant part of research conducted under the label of "Anthropocene" in either the geosciences or social sciences. The strongest geoscience keywords (Table 3), for example, are related to sedimentation and hydrology, indicating an almost total

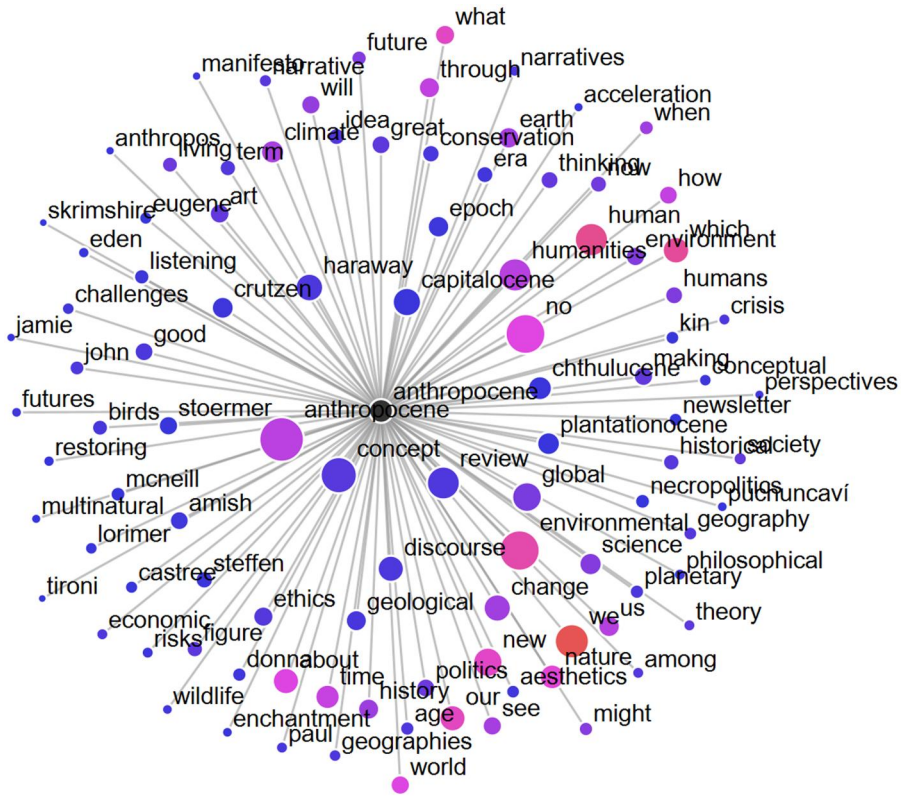

Figure 2. Humanities Collocates

lack of research relating to those processes in the social sciences. And topic modeling (Table 5) shows that research is highly segregated, with geologic processes studied almost exclusively in the earth sciences, and social processes are almost entirely absent in those same sciences. We found this result surprising, given that the whole idea of the Anthropocene is that social and earth system processes have blended. For this reason, we expected to see a significant amount of interdisciplinary work about the interaction of human and natural processes in a corpus of research on the Anthropocene. But the data indicate that while recognition that human and geologic systems interact is widespread, serious investigation into that interaction is vanishingly rare. This result is especially keen given that our corpora contain only those articles containing the term *Anthropocene*, and so represent the articles most likely to contain research on human–geologic interaction.

Even some areas of apparent overlap reveal divides upon close examination. Both the Geoscience and Humanities corpora, for instance, feature many keywords and collocate dealing with space and time.[5] But keywords (Tables 1 and 2) demonstrate that space and time mean different things between the two disciplinary families. Geologists treat these as objective, measurable concepts (see keywords like *bp, ka*, and *dating*). Humanities scholars, on the other hand, are concerned with the subjective experience of space and time, as shown by their distinctive concern with the topic "Phenomenology of Nature," for instance. If the shared occupation with questions of space and time is to serve as a bridge facilitating cross-discipline collaboration about the Anthropocene, these differing conceptualizations of space and time will need to be at least partially reconciled. Some scholars have initiated efforts to do so (e.g. Chakrabarty 2018), but much more work in this area will need to be done if space and time are going to be an inroad between disciplines.

Similarly, topic and collocate data show that Humanities and Social Science corpora demonstrate a shared interest in people and society (Table 5 and Figs 2 and 3). That is not surprising. But looking at keywords shows that this surface commonality covers a deeper division. Compare the personal pronouns dominant in the humanities (Table 2) to the interest in *institutions* and *systems* in the social sciences (Table 4 and Fig. 3). This division between individualistic and holistic perspectives on people and society would need to be overcome for the humanities and social sciences to engage in shared research on the Anthropocene.

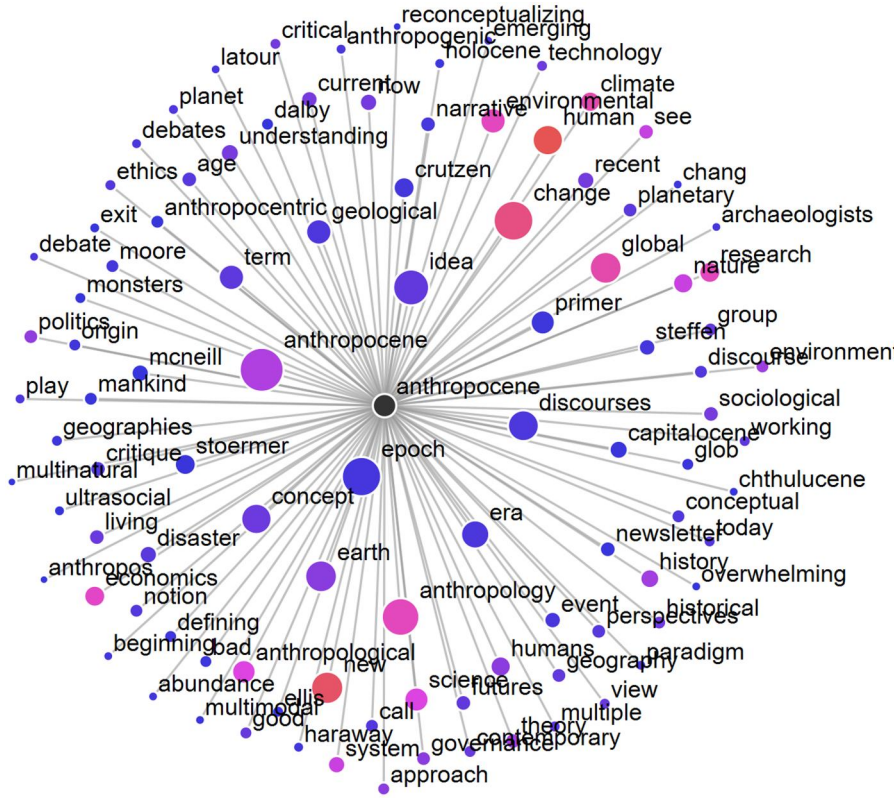

Figure 3. Social Science Collocates

In corpora about the Anthropocene, we would expect, at the most basic level, common interest in global environmental change. Indeed, collocate data indicate that all three corpora discuss the Anthropocene in the same breath as climate change and other forms of environmental change (Figs 1–3). But topic modeling shows that what types of change are of interest to each discipline differs; earth scientists are much more interested in climate change than the other disciplines, for instance (Table 5). Analysis of keywords highlights the shocking extent to which this is the case: CO_2 is a strong keyword for the geosciences compared to both the humanities (Table 1, SMP 7.08) and the social sciences (Table 3, SMP 5.12), and other terms central to climate change, such as *temperature* and *methane* are also distinctive to geology. These are not highly technical natural scientific terms—CO_2 and *temperature* are terms we would expect in any detailed research about climate change—so the overwhelming difference between corpora in their frequency suggests a lack of engagement with the topic in the social sciences and humanities. Since climate change is the most discussed Earth System process of the Anthropocene, this gap indicates a fundamental disconnect between what geologists mean by *Anthropocene* and what others do.

An even deeper disconnect, in our view, is that the humanities and social sciences examine global change through the lens of value, as brought out by collocates and keywords like *ethics, justice, moral, welfare*, *care, crisis*, and *disaster* (Tables 2 and 4; Figs 2 and 3). Value-laden terminology is almost entirely absent from the Geoscience corpus, by contrast, suggesting an intent—to be expected from natural science—to study global change through a purely descriptive lens. The fact that the natural sciences aim to be value-free,[6] while other disciplines do not does not entail that the various disciplines cannot inform each other. But it does problematize the possibility of a unified study of the Anthropocene. Should interdisciplinary Anthropocene studies be value-driven or strictly descriptive? If humanities scholars and social scientists participate in the process of formalizing the Anthropocene as a unit of the GTS, should they bring their ethical and political stances with them? These sorts of questions need to be settled if collaborations between geologists and other researchers are to be successful.

The corpus analysis, as we read it, presents bleak prospects for comprehensive interdisciplinary Anthropocene studies. It also, however, points to a number of narrower areas of existing or potential cross-disciplinary work around the Anthropocene, such as the shared interest

Table 5. Average topic weights

Topics	Geoscience	Humanities	Social Science
Volcanism	0.0173	0.0060	0.0025
Literature	0.0014	0.1664	0.0066
Contemporary Climate Change	0.1462	0.0082	0.0088
Ocean geology	0.0943	0.0015	0.0028
Phenomenology of nature	0.0105	0.1644	0.0384
Anthropology	0.0282	0.0523	0.1049
Environmental Policy	0.0432	0.0246	0.1395
GIS	0.1461	0.0188	0.1073
Fluvial sedimentation	0.0476	0.0015	0.0015
Non-human animals	0.0007	0.0789	0.0285
Species Conservation	0.0084	0.0176	0.0276
Social Justice	0.0022	0.0431	0.0830
Economics	0.0061	0.0043	0.1093
Dating	0.0910	0.0014	0.0013
China	0.0536	0.0055	0.0089
Freshwater Hydrology	0.0515	0.0016	0.0061
Humans and society	0.1088	0.3582	0.2664
Lacustrine Geology	0.0511	0.0005	0.0006
Landscape archaeology	0.0903	0.0040	0.0192
Indigenous studies	0.0015	0.0414	0.0367

Cells are saturated (i.e. redder) in proportion to topic weight across each corpus.

between geoscience and social science in GIS methods (Table 5), in which there is already much collaborative work. We also noticed a shared recognition between humanities scholars and geoscientists that the Anthropocene is a philosophical question (Figs 1 and 2), suggesting the potential for fruitful collaboration between geologists and philosophers of science. Other areas of promising overlap include the humanities and social sciences both connecting the Anthropocene to indigenous studies (Table 5) and how researchers in the humanities and social sciences are concerned with how the Anthropocene relates to non-human animals and species conservation (Table 5).

In short, the results of our corpus analysis indicate that the best opportunities for interdisciplinary collaboration surrounding the Anthropocene are probably focused and narrow questions of these sorts. Shared interests in techniques (e.g. GIS) or in "the Anthropocene and *x*" (e.g. indigenous communities, species conservation) are common and facilitate collaborative work. These are narrow bands of connection within a broad channel of difference, however. Without radical disciplinary change, our data show much stiffer barriers to the blurring of disciplinary boundaries to tackle the big questions about anthropogenic global change, and to bring on board social scientists and humanities scholars in addressing formal issues of the GTS. Given the degree of separation between what each discipline takes the Anthropocene to be as an object of study, deep collaborative research seems unlikely.

5. Conclusion: how are attempts at tackling the Anthropocene across the disciplines faring?

The data as we read them suggest a real potential for interdisciplinary work on the Anthropocene, but mostly where that work has a narrow, specific focus. We are much less confident that we can achieve the pan-disciplinary approaches the Anthropocene scholars have been calling for (e.g. Chakrabarty 2009; Castree 2014; Thomas 2014; Ellis *et al.* 2016; Lock 2018). This is not to say that research involving all three families of disciplines cannot be valuable, merely that it cannot address overarching shared issues, such as dating the geological onset of the Anthropocene or developing a strategy for tackling the climate crisis—because these issues aren't shared topics of concern across groups of disciplines.

We come to that conclusion based on the academic corpus, but a review of existing attempts to bridge natural and social sciences and the humanities illustrate the same point. For example, *The Anthropocene: A Multidisciplinary Approach* (Thomas et al. 2020) canvasses the approaches to the Anthropocene from several disciplines: geology, Earth System science, biology, history, anthropology, and economics. But Thomas *et al.* have a good understanding of the differences in methods and concepts between these disciplines, and so do not really try to present a unified research program. For the most part, each chapter of the book explains the current state of research on the Anthropocene within a particular discipline, and the

goal of the book is to begin creating "networks of data and stories" rather than aiming for integration. As a proof of concept of such a network, the book is successful (Morehouse 2021), but the network it creates is a loose, fragmented one. The book highlights potential connections between the disciplines, but those connections are not the sort that could serve as the foundation of a unified approach to tackling any of the big questions about the Anthropocene.

Another book drawing on geology, social science, and the humanities ends up in a very different place, but illustrates the same point. *A billion black Anthropocenes or none* (Yusoff 2018) flags itself as "transdisciplinary" and draws on geology, literature, and social science. On the surface, it would appear to be of a kind to something like *The Anthropocene: A Multidisciplinary Approach*, since it purports to tackle some of the same issues, such as determining the geological boundary of the Anthropocene, or the politics of the environmental crisis. But the resemblance is only skin deep. Yusoff, for instance, argues that "geology is a hinge that joins indigenous genocide, slavery, and settler colonialism" and she is primarily concerned with how our ways of thinking about the planet might reinforce social injustices. These are topics almost totally absent from Thomas et al. (2020), and, conversely, the data and stories in Thomas *et al.*'s "network" fail to appear in Yusoff's book. Our contention is that this is not an accident or a failure on the part of one of the two books. Instead, to think that both Thomas *et al.* and Yusoff are interested in overlapping issues is to be misled by the fact that they both claim to be writing about the "Anthropocene." What that label means to each set of authors is largely distinct, and the pair of books might both be valuable research, but they are not studies that speak to each other in any deep way. Nor do they need to be: if we are right, the concept *Anthropocene* has speciated, and each of these purportedly multidisciplinary books tackles a different species of "Anthropocene."

What does this mean for calls to, say, get humanists and social scientists significantly involved in revising the GTS (Ellis *et al.* 2016), or to dissolve the boundaries between history and geology (Chakrabarty 2009)? If we are reading the data right, these are not going to happen anytime soon. This is not to say that researchers from around the academy should not continue to collaborate and inform each other, but to recognize that fruitful collaboration is most likely to occur in well-defined areas of mutual interest, rather than by trying to use the wide-ranging concept *Anthropocene* to rig together cross-disciplinary bridges.

Supplementary data

Supplementary data is available at *DSH* online.

Author contributions

Carlos G. Santana (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing), Kathryn Petrozzo (Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Validation, Writing—original draft, Writing—review & editing), and Timothy Perkins (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing).

Funding

This research was supported by the Digital Matters program at the University of Utah.

Notes

1. Leading journals in other social science fields, such as political science and linguistics, had too few articles mentioning the Anthropocene to be worth including.
2. Geosciences: 1,518,585 words, 163 articles
   Humanities: 1,457,848 words 213 articles
   Social sciences: 1,501,188 words, 147 articles
3. Thanks to our anonymous referees for pressing us to elaborate on this point.
4. Why twenty? LDA will always generate, however, many topics you assign it to, so to some extent the number of topics is arbitrary. But assigning too few topics yields topics so broad to be uninformative. Assign too many, and your algorithm will meet the demand by reading structure into the noise, which means spurious results. We chose twenty as a likely bet to navigate between those perils.
   We could have generated topics at a variety of numbers and selected the "best" set, but that would have introduced the ability for us to unconsciously select the set of results that best fit our biases and prejudices. Instead, we determined from the get-go to work with whatever result twenty topics provided us. Our bet paid off—we think that if you look at the set of topics we generated, you will agree that they are clear, intuitive, and illuminating.
5. Puzzlingly, the Social Science corpus does not focus on space and time to the same extent, despite containing papers from two geography journals.
6. Our corpus data are evidence for the unsurprising claim that value-neutrality is a standard in geoscientific discourse. That claim does not entail that these sciences are *in fact* value neutral or value free. They probably are not (Havstad and Brown 2017), and philosophers of science have long argued that such neutrality is not necessary for scientific objectivity (Longino 2002). Nevertheless, the fact that a standard of value-neutrality is stronger in the geosciences than in the social sciences or humanities does present a barrier to transdisciplinarity.

References

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003) 'Latent Dirichlet Allocation', *Journal of machine Learning Research*, **3**: 993–1022.

Brezina, V. and Platt, W. (2023) *LancsBox X [software]*. England: Lancaster University.
Brezina, V., Timperley, M., and McEnery, T. (2018) '*LancsBox v. 4. x [software]*'. https://lancsbox.lancs.ac.uk/.
Castree, N. (2014) 'The Anthropocene and the Environmental Humanities: Extending the Conversation', *Environmental Humanities*, **5**: 233–260.
Chakrabarty, D. (2009) 'The Climate of History: Four Theses', *Critical Inquiry*, **35**: 197–222.
Chakrabarty, D. (2018) 'Anthropocene Time', *History and Theory*, **57**: 5–32.
Chang, J. *et al.* (2009) 'Reading Tea Leaves: How Humans Interpret Topic Models', *Advances in Neural Information Processing Systems* **22**, 288–96. https://proceedings.neurips.cc/paper/2009.
Crutzen, P. J. and Stoermer, E. F. (2000) 'The "Anthropocene"', *Global Change Newsletter*, **41**: 17–8.
Ellis, E. *et al.* (2016) 'Involve Social Scientists in Defining the Anthropocene', *Nature*, **540**: 192–93.
Haraway, D. (2015) 'Anthropocene, Capitalocene, Plantationocene, Chthulucene: Making Kin', *Environmental Humanities*, **6**: 159–65.
Havstad, J. C. and Brown, M. J. (2017) 'Neutrality, Relevance, Prescription, and the IPCC', *Public Affairs Quarterly*, **31**: 303–24.
Hayes, C. T. *et al.* (2017) 'Helium and Thorium Isotope Constraints on African Dust Transport to the Bahamas Over Recent Millennia', *Earth and Planetary Science Letters*, **457**: 385–94.
Hicks, D. J. (2021) 'Productivity and Interdisciplinary Impacts of Organized Research Units', *Quantitative Science Studies*, **2**: 990–1022.
Kilgarriff, A. (1997) 'Using word frequency lists to measure corpus homogeneity and similarity between corpora', In: *Proceedings of the Fifth Workshop on Very Large Corpora*. Beijing: Tsinghua University.
Kilgarriff, A. (2009) 'Simple Maths for Keywords', In: *Proceedings of the Corpus Linguistics Conference*. University of Liverpool. https://ucrel.lancs.ac.uk/publications/cl2009/.
Lock, M. (2018) 'Mutable Environments and Permeable Human Bodies', *Journal of the Royal Anthropological Institute*, **24**: 449–74.
Longino, H. E. (2002) *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.
McCallum, A. K. (2002) '*Mallet: A Machine Learning for Language Toolkit' [software]*, http://mallet.cs.umass.edu, accessed 19 Oct. 2019.
Mizrahi, M. and Dickinson, M. A. (2022) 'Is Philosophy Exceptional? A Corpus-Based, Quantitative Study', *Social Epistemology*, 37:5,666-683
Morehouse, H. (2021) 'Book Review: The Anthropocene: A Multidisciplinary Approach', *Environmental Politics*, **30**(4): 682–684. doi: 10.1080/09644016.2021.1892120.
Pence, C. H. and Ramsey, G. (2018) 'How to Do Digital Philosophy of Science', *Philosophy of Science*, **85**: 930–41.
Rychlý, P. (2008) 'A lexicographer-friendly association score', In: *Proceedings of Recent Advances in Slavonic Natural Language Processing* (pp. 6–9). Masaryk University. http://nlp.fi.muni.cz/raslan/2008/.
Thomas, J. A. (2014) 'History and Biology in the Anthropocene: Problems of Scale, Problems of Value', *The American Historical Review*, **119**: 1587–607.
Thomas, J. A., Williams, M., and Zalasiewicz, J. (2020) *The Anthropocene: A Multidisciplinary Approach*. Hoboken, NJ: John Wiley & Sons.
Yusoff, K. (2018) *A Billion Black Anthropocenes or None*. Minneapolis, MN: University of Minnesota Press.
Zalasiewicz, J. *et al.* eds. (2019) *The Anthropocene as a Geological Time Unit: A Guide to the Scientific Evidence and Current Debate*. Cambridge: Cambridge University Press.

Appendix 1—list of journals

Geosciences:
Earth and Planetary Science Letters
Earth-Science Reviews
Geosphere
Journal of Quaternary Science
Nature Geoscience
Quaternary Geochronology
The Holocene

Social Sciences:
American Anthropologist
Applied Geography
Ecological Economics
Journal of the Royal Anthropological Institute
The Geographical Journal
The Sociological Review

Humanities:
Environmental Humanities
GeoHumanities
Interdisciplinary Studies in Literature and Environment
Modern Language Quarterly

Digital Scholarship in the Humanities, 2024, 39, 723–735
https://doi.org/10.1093/llc/fqae004
Full Paper