

Race Science in Mainstream Psychology, 1960-2010

Abstract

While anthropologists and human geneticists have long rejected the biological race concept, both social movements and prominent researchers have purported to justify racial inequality and colonialism by appealing to the epistemic authority of science.

In this study, bibliometric and text mining methods were used to trace the appearance of race science discourse in both mainstream behavioral science journals and the parascholarly journal *Mankind Quarterly* from 1960 to 2010. Our analysis found two distinguishable lines of race science discourse. One is strongly associated with *Mankind Quarterly*, while the other — emphasizing race and intelligence — is strongly associated with mainstream journals, especially *Intelligence* and *Personality and Individual Differences*. Psychology, and especially intelligence research, has played a significant role in mainstreaming race science.

By the end of the Second World War, a combination of technical developments in anthropology and genetics, along with revelations of Nazi atrocities, had led most members of the scientific community to reject attempts to scientifically justify racial inequalities and colonialism (1). But eugenic thinking and scientific racism persisted throughout the second half of the twentieth century, and recent work has found that white supremacists continue to attempt to use scientific research to justify racial violence (2–4).

The aim of the study reported here was to trace race science in the scholarly and parascholarly literature, examining the ways that fringe ideas can present themselves in mainstream venues as legitimate scholarly inquiry.

We first propose a distinction between scientific racism, race science, and race science discourse. *Scientific racism* refers to the social practice of purporting to justify racial inequality and colonialism by appealing to the epistemic authority of science. *Race science* refers to scientific and/or pseudoscientific research, including research products (journal articles, etc.), that can be utilized for scientific racism. *Race science discourse* refers to a broader category, any treatment of race science as a legitimate area of scientific research. This includes methodological and empirical critiques of race science. For example, the Flynn effect

*Corresponding author

(the finding that IQ scores have increased over time) poses a serious empirical challenge to hereditarian claims of racial differences in intelligence. But technical debates over the extent and causes of the Flynn effect likely create the impression among the general public of ordinary scientific disagreement. In this way technical critiques of race science based on the Flynn effect could potentially legitimize race science. So empirical research supporting the Flynn effect is usually not itself race science, but can still be race science discourse.

These distinction allow us to bracket the intentions and attitudes of the researchers involved in any particular piece of research or debate. For the purposes of the current study, it is not important whether, for instance, psychologist Arthur Jensen is or is not correctly labelled a scientific racist (5). But it is important that much of Jensen's research was useful and indeed utilized to purportedly justify racial inequality, and so unambiguously qualifies as race science.

Histories of scientific racism in the twentieth century have often emphasized the Pioneer Fund and the parascholarly journal *Mankind Quarterly* (MQ) (1,6–13). Pioneer was formed in 1937, during the waning years of the eugenics movement in North America (1), with the aim of “support[ing] academic research and the '[sic]dissemination of information, into the ‘problem of heredity and eugenics’ and ‘the problems of race betterment’” (6, quoting Laughlin). MQ was founded in 1960 by biologist R. Ruggles Gates (1882–1962), psychologist Henry Garrett (1894–1973), and non-academic anthropologist G. Robert Gayre (self-styled as “Gayre of Gayre and Nigg”; 1907–1996). Gates' professional status had risen and fallen with eugenics and the explicit scientific racism of the 1920s, and by the end of the Second World War he was thoroughly marginalized (7). In contrast, Garrett had been president of the American Psychological Association in 1946 and chair of Psychology at Columbia from 1941 to 1955 (7).

In the landmark case *Brown v Board of Education of Topeka* (1954), the US Supreme Court banned *de jure* educational segregation. The Court's decision relied on expert testimony from psychologists and education researchers; but the segregationists also put forward their own experts, including Henry Garrett (7,9,14). As part of this segregationist reaction, MQ was created to provide a favorable venue for race scientists to publish their views, on the grounds that an “equalitarian dogma” created a censorious “taboo” against their research in mainstream publications (8,15).

In light of the scientific racist origins of Pioneer and MQ, and the significant attention MQ has received in recent historiography, we hypothesized that bibliometric and text mining analyses would show race science ideas originating in MQ, and from there being disseminated to mainstream publications. This hypothesis was not supported. Instead, mainstream psychology journals provided a venue for a distinct form of race science, based on intelligence research and originating outside of MQ.

Results

Mankind Quarterly and Pioneer-funded researchers

We identified 16 researchers who had received funding from Pioneer; 14 of these researchers had profiles in the Web of Science (WoS) author search, allowing us to identify 13 WoS-indexed journals that had published 6 or more of these authors. See Table 1 and figure 1.

Table 1: Pioneer-funded researchers. Either identified in (16) or named on an archive copy of Pioneer’s website, along with birth and death dates from Wikipedia, attributed discipline, and WoS author search result counts. Brunetto Chiarelli does not have a Wikipedia page.

Thomas J. Bouchard, Jr. (1937-)	psychology	184
Brunetto Chiarelli (?-?)	anthropology	91
Hans Eysenck (1916-1997)	psychology	661
Robert A. Gordon (1932-)	sociology	0
Linda Gottfredson (1947-)	psychology	71
Garrett Hardin (1915-2003)	ecology	75
Joseph M. Horn (1940-)	psychology	68
Lloyd Humphreys (1913-2003)	psychology	0
Arthur Jensen (1923-2003)	psychology	235
Michael Levin (1943-)	philosophy	96
Richard Lynn (1930-2023)	psychology	288
R. Travis Osborne (1913-2013)	psychology	59
J. Phillippe Rushton (1943-2012)	psychology	277
Audrey M. Shuey (1900-1977)	psychology	10
Philip A. Vernon (1950-)	psychology	227
Daniel Vining, Jr. (1944-)	demography	33

Figure 1 shows that, while MQ is among the “Pioneer-publishing” journals, a number of mainstream journals are more prominent: *Personality and Individual Differences* (PID), *Intelligence* (Int), *Behavior Genetics* (BG), and *Psychological Reports* (PR). In addition, only psychologist Richard Lynn appears to have published heavily in MQ. Lynn became an assistant editor of MQ in 1979 (vol. XX, Nos. 1 & 2) and is listed as editor-in-chief on MQ’s current website as of 2023-07-21. He was also president of Pioneer from the death of psychologist J. Phillippe Rushton in 2012 (17) until his own death in 2023. By contrast, a number of Pioneer-funded researchers have published a significant number of papers in PID, Int, and to a lesser degree BG: Bouchard, Eysenck, Jensen, Rushton, Vernon, and also Lynn.

Personality and Individual Differences (PID) was founded in 1980, with Eysenck as editor-in-chief and an editorial board including Jensen and Lynn. In the inaugural editorial, Eysenck identified “studies of the genetic determinants of individual differences in the areas of personality and intelligence” as one of the journal’s eight major areas of interest. Eysenck remained editor-in-chief

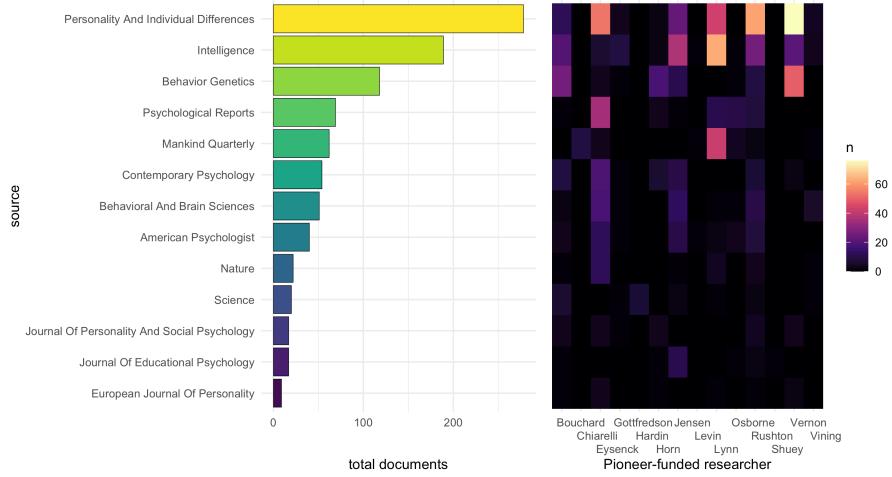


Figure 1: Journals publishing 6 or more Pioneer-funded researchers, Web of Science author search results.

until his death in 1997. In 2005 the editorial board still included Jensen and Lynn. PID was first published by Pergamon Press, a mainstream academic press, and today is published by Elsevier.

Intelligence (Int) was founded in 1977, with psychologist Douglas Detterman as editor-in-chief from the founding until 2016. Lloyd Humphreys was on the editorial board starting from 1977; by 1990 he had been joined by Jensen and Philip Vernon. Richard Lynn joined the editorial board sometime between 1998 and 2002. (Archive copies of the Int editorial board page are not available from the journal’s website from 1999 through 2001.) Int has been criticized for including Lynn and Gerhard Meisenberg — who was editor-in-chief of MQ in 2015-18 — on its editorial board until 2018 (10). Int is published by Elsevier.

Topic model analysis identifies race science discourse

The fact that Pioneer-funded researchers published heavily in two mainstream psychology journals does not tell us anything about the content of their publications or the claims of their research. We assembled a full-text corpus of articles published in MQ, PID, Int, BG, PR, and Behavior and Brain Sciences (BBS) between 1960-2010 and used topic modeling to identify topics discussing race, intelligence, and both. (See the materials and methods for a discussion of some important limitations with this corpus: journals published by the American Psychological Association could not be included, BBS bundles focus articles and commentaries together in a way that confounds simple document counts, and an unknown number of relevant documents in PID could not be included.)

Figure 2 and this discussion focuses on three topics identified in one of the 24 fitted topic models; see figures ED-1, ED-2, ED-3, ED-4, ED-5, ED-6 for all

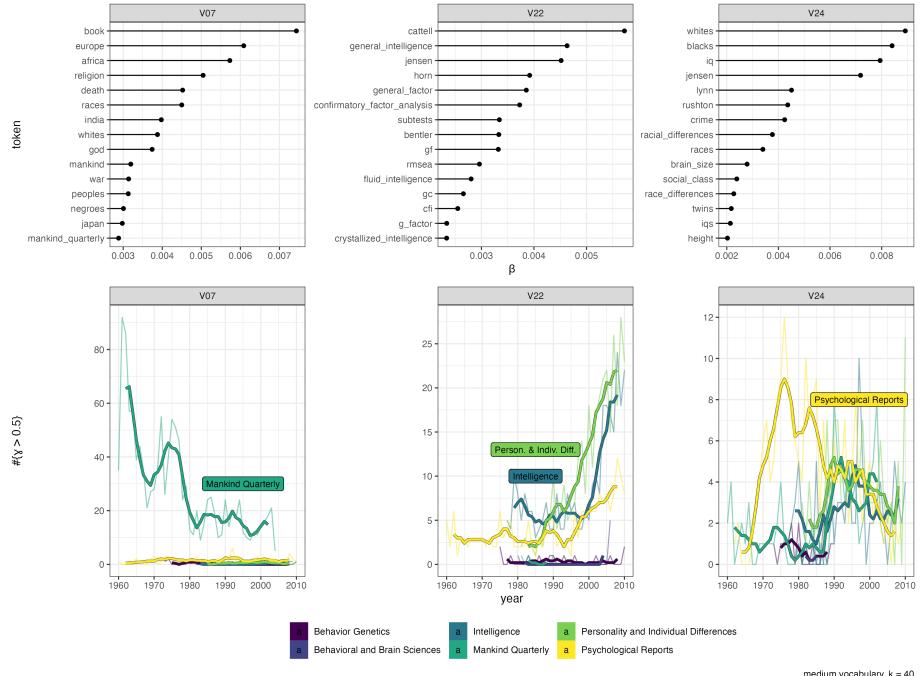


Figure 2: Silge plots (18) and smoothed time series for three focal topics. Top row: top 15 terms by β (term-topic distribution) for each topic. Bottom row: count of articles associated with each topic, by journal and year. Article counts use a threshold approach, with γ (topic-document distribution) greater than 0.5. Thin lines give annual values, thick lines give 5-year running averages. Because *Behavior Genetics* and *Behavioral and Brain Sciences* are not prominent in any panel, neither is given a direct label.

topics from all models, and figures ED-7, ED-8, ED-9 for selected topics from all models. In the focal model, topic 07 is strongly associated with MQ: MQ published dozens of articles in this topic each year, and no other journal ever published more than a handful. The “Silge plot,” showing the top 15 terms in the topic (18), contains a number of racial terms (**races**, **whites**, **negroes** and potentially **europe**, **africa**, **india**, and **japan**); as well as **book**, likely reflecting the fact that MQ published a number of book reviews, while the mainstream journals either did not publish book reviews or these were not available for the corpus.

Topic 22 is strongly associated with Int and PID in the same way that 07 is associated with MQ. The Silge plot does contain **jensen**, as well as a reference to Raymond Cattell, who played a major role in the development of factor analysis and intelligence testing but also advocated for eugenics, fascism, and Nazi race science (19). However, the other authors named in this topic — John Horn and Peter Bentler — do not appear to have contributed to scientific racism. Instead this topic appears to identify “mainstream” (non-race science) intelligence research, especially factor analysis and the debate over whether intelligence is unidimensional or multidimensional. (Three other topics only associated with mainstream intelligence journals and terms were also identified by this topic model.) This topic indicates that the model is able to distinguish “mainstream” intelligence research from race science discourse.

The Silge plot for topic 24, by contrast, suggests a distinct race science discourse topic, with multiple racial terms (**whites**, **blacks**, **racial_differences**, **races**, **race_differences**) and the names of three prominent Pioneer-funded researchers, **jensen**, **lynn**, and **rushton**. Independent qualitative coding of 121 papers in this topic (those with $\gamma > 0.97$ and published in journals other than PR) confirmed this interpretation, with 108 (89%) coded as race science discourse by both authors (Cohen’s $\kappa = 0.86$).

In almost all years, most papers in topic 24 were published in mainstream journals rather than MQ. Jensen’s “How Much Can We Boost IQ and Scholastic Achievement?” (20) was published in 1969 in *Harvard Educational Review* (not included in this study), and the time series indicates that, in the early 1970s, there was an increase in articles in topic 24 in both PR and MQ (the only two journals in our corpus that were active at the time). MQ shows another sharp increase in the late 1980s; using content analysis, Adams and Pilloud found that psychology was the dominant discipline in MQ in the period 1992-2018 (12). PID published multiple papers in topic 24 almost immediately after it was founded, with Int showing a more gradual increase between the mid-1970s and mid-1990s.

Race science, behavior genetics, and psychology

BG published very few articles in any race-and-intelligence topic identified by any of the 24 topic models. Among both academics and the general public, the

field of behavior genetics is strongly associated with race science, and specifically race-and-intelligence research. Panofsky (21) argues that, prior to Jensen's 1969 paper, behavior genetics emphasized the study of non-human animals and intentionally avoided associations with eugenics and public controversy more generally. In response to Jensen, critics such as Lewontin offered broad critiques of behavior genetics as such, and behavior geneticists in turn adopted a radical conception of academic freedom (without any sense of responsibility for the social implications of academic research) and a siege or wartime mentality, as illustrated by Sandra Scarr's 1986 presidential address to the Behavior Genetics Association (BGA) (22). Scarr's address coincided with the period between 1970-1990 when BG published articles in topic 24 (including Scarr's address itself).

However, the topic model analysis suggests that, by the 1990s, behavior genetics may have distanced itself somewhat from race science, albeit without directly repudiating it. This interpretation was supported by a supplemental analysis that focused on BG specifically (see methods and materials). In addition, the 1995 BGA presidential address by Glayde Whitney — in which he criticized the “Marxist-Lysenkoist denial of genetics” and proposed that differences in murder rates between countries and cities were caused by racial genetic differences in intelligence, empathy, aggression, and impulsivity — was published in MQ rather than BG. Behavior genetics may have been prominent in promoting race science in the past, but appears to have been less receptive to such ideas in more recent history. Instead, our results suggest that other subdisciplines of psychology have provided the primary venue for mainstreaming race science.

Discussion

Using bibliometric and text-mining methods, this study finds that Pioneer-funded researchers published heavily in certain mainstream psychology journals, much more than in *Mankind Quarterly*; and that a distinct topic of race science discourse, centered on race and intelligence, can be identified in these same mainstream journals across the period 1960-2010. These findings indicate that *Mankind Quarterly* (MQ) was less important as venue for late 20th century race science than mainstream psychology journals, especially *Intelligence* (Int) and *Personality and Individual Differences* (PID). Indeed, until the 1990s, MQ published a very different kind of race science from the race-and-intelligence research published in Int and PID. During this same period of time, race scientists served in prominent and influential positions in behavior genetics and psychology, including as society presidents and members of journal editorial boards. In some cases, some race scientists simultaneously maintained active connections to both the mainstream scientific community and white supremacist organizations (5,11,23).

We are not identifying any particular scientist as a racist, scientific racist, or race scientist based on the topic model results alone. Such claims require an analysis of documentary evidence that goes beyond the scope of the current study

(5). We do assume that our readers, like us, regard white supremacy and scientific racism as morally odious and beyond the scope of reasonable debate (24). However, scientific research can be appropriated to promote scientific racism — and thus count as race science — even when this is contrary to the intentions of the original researchers themselves (25–28). Panofsky et al. show how the term “human biodiversity,” originally developed for anti-racist purposes by biological anthropologist Jonathan Marks, has been inverted by white supremacists and is now a dogwhistle for biological racial hierarchy (3).

Over the past few decades, many professional organizations in genetics and anthropology have made formal statements rejecting race as a biologically meaningful concept (29–36), undermining the core assumption of race science. The American Psychological Association (APA) — whose members are predominantly clinical psychologists — the Federation of Associations in Behavioral & Brain Sciences (FABBS) — an organization whose purpose is to provide policy recommendations — and the editors of *Nature Human Behavior* have openly rejected race as biologically meaningful (37–39). But research-focused psychological organizations like the Association for Psychological Science (APS) and Psychonomic Society (PS) stop short of rejecting race as biologically meaningful when they denounce racism (40,41). Psychology as a discipline maintains space for several mythological race science narratives, such as allegations of a taboo against race and intelligence research (42), claims that scientists engaging in or calling for any kind of anti-racism in the field is an ideological corruption of dispassionate and value-neutral science (43), or arguments that holding race science to the same evidentiary standards as other psychological research is a violation of academic freedom (44). And historically several race scientists have secured important gatekeeping positions within the scientific community, such as on the editorial boards of *Intelligence* and *Personality and Individual Differences*.

Research on the genetics of intelligence has been socially harmful, not just in the somewhat abstract sense of promoting racial stigma, but in the concrete sense of being used to rationalize mass shootings and other acts of racial violence (4,45). While scientists do not fully control the downstream social effects of their research, like other citizens scientists are responsible for mitigating reasonably foreseeable harms that result from their actions (46–49). These actions include not only individual decisions to research this topic or that, but also collective decisions about the allocation of scarce resources such as journal space, research funding, and positions of power and influence. Given the way research on intelligence and behavior genetics has been used historically, these fields may be especially susceptible to appropriation by scientific racists. This only strengthens the obligations of researchers in these areas to prevent their research from causing harm (27,50).

How should psychologists exercise these responsibilities? We believe restorative or reparative justice provides a useful model (51,52). Unlike retributive justice (used by almost all criminal justice systems today), the aim of restorative

justice is to repair the damaged relationships between victims and perpetrators of injustice. Typically, restorative justice requires perpetrators to not only acknowledge their actions and the harms that these actions had on victims, but also to work with victims to identify concrete actions that perpetrators can take to redress or mitigate these harms. From this perspective, explicit apologies for contributions to scientific racism are essential as a first step (33,37,53), but not fully sufficient on their own.

Materials and methods

Corpus assembly

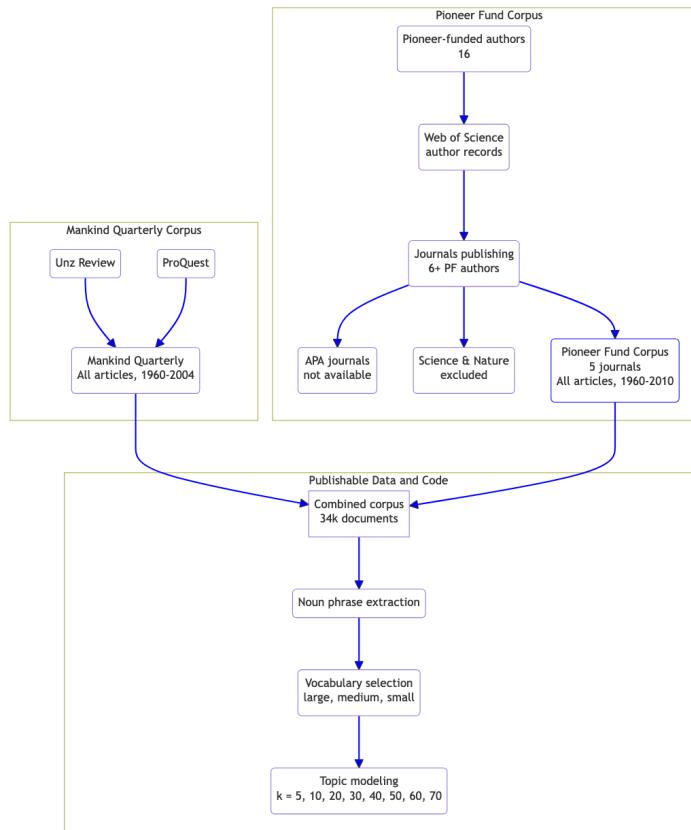


Figure 3: Corpus assembly.

Corpus assembly is summarized in figure 3. The corpus was assembled in two parts, one for *Mankind Quarterly* and the other for mainstream journals.

Mankind Quarterly

An electronic archive of *Mankind Quarterly*, from the first issue in 1969 through 2004, is available for free on the open web at the white nationalist website *The*

Unz Review. The Python package Beautiful Soup (<https://www.crummy.com/software/BeautifulSoup/>, version not recorded) was used to retrieve every PDF available in this archive. After identifying some gaps (missing issues) in the archive from *The Unz Review*, we manually retrieved PDFs for further articles using ProQuest. Text was extracted using the R package `pdftools` (54 version 3.0.1). This stage of corpus assembly was conducted in Fall 2021 with the assistance of Anthony Sainez.

Mainstream journals

To identify suitable mainstream journals for inclusion in the corpus, we first identified academic researchers who had been funded by the Pioneer Fund. Pioneer is an American non-profit organization founded in 1937 to fund eugenics research and propaganda (8). After the *Brown v. Board of Education* ruling in 1954, Pioneer funded various segregationist efforts across the United States, including lectures on eugenics by Stanford physicist William Shockley (14). We reviewed a critical profile of Pioneer (16) as well as an archived page from the organization's own web site (<https://web.archive.org/web/20130103005545/http://www.pioneerfund.org/Grantees.html>), which together listed 16 researchers who had received Pioneer funds.

We then used the author search tool in Clarivate's Web of Science platform (<https://www.webofscience.com/wos/author/search>), retrieving publication lists for 14 researchers. These searches were conducted between 2021-09-24 and 2021-10-05 by [AUTHOR 1]. After parsing these results, we counted how many of the 14 researchers had published in each journal. Thirteen journals had published 6 or more of the 14 Pioneer-funded researchers. We excluded *Mankind Quarterly* (as already included) as well as *Science* and *Nature* (as too general) from further consideration in this side of the corpus. Three journals published by the American Psychological Association (APA; *American Psychologist*, *Contemporary Psychology*, and *Journal of Educational Psychology*) had to be excluded due to confusion over who could give us permission to use the archives for a text mining project, with both APA and ProQuest asserting that we needed to get permission from the other entity. *European Journal of Personality*, published by SAGE, also had to be excluded because our institutional access only went back to 1999.

The remaining 5 journals are all published by major academic publishers — Elsevier, Springer, or Cambridge University Press — and each item in the entire run of each journal has been assigned a DOI (digital object identifier) for archival purposes. We used the Crossref API and `rcrossref` R interface to this API (55 version 1.1.0.99) to retrieve metadata for each item published from 1960-2010 in each of these journals. These metadata included item-level license information — confirming that the text of each item could be used for text mining projects — and a URL to an electronic version of the item. These URLs were used to retrieve a XML or PDF version of each item, except for *Personality and Individual Differences*. This journal has published a relatively large number of non-article documents, such as book reviews and commentaries, that are not

available at the URL included in the Crossref metadata. (This is unfortunate, as it was not difficult to find highly relevant documents that we could not automatically retrieve and therefore had to be excluded from our corpus. One set of examples is an exchange between Rushton and Flynn on the Flynn effect (56–58).) Instead we used Elsevier’s ScienceDirect API (<https://dev.elsevier.com/>) to independently search and retrieve all available items from *Personality and Individual Differences*. This stage of corpus assembly was conducted between 2021-11 and 2022-05.

Behavioral and Brain Sciences typically uses a target article + commentary format; in some cases there are dozens of commentaries for a single article. In the Crossref DOI metadata, each target article and individual commentary is given its own DOI, with no distinction between contribution types or metadata links between a commentary and the corresponding target article. But text is only available in aggregate PDFs that bundle together the target article and all of the commentaries. In some cases this results in PDFs that are hundreds of pages long and might be linked from the Crossref metadata 30+ times. In addition, the retrieved PDFs are not perfectly identical, because Cambridge UP’s servers add a timestamped watermark to each page when the PDF is requested. We attempted to contact Cambridge for assistance but did not receive a response. We ultimately used a series of ad hoc measures to mitigate text duplication. Approximately 100,000 documents that were 1-2 pages long (per Crossref metadata) were excluded before PDF retrieval. After PDF retrieval and text extraction, the timestamped watermarks were removed using a regular expression and the text was hashed using SHA 256. Hashes were used to efficiently construct groups of identical documents, and a single document (whichever one happened to be first in the dataframe) was chosen from each hash group for inclusion in the corpus. It is plausible that some duplicates made it through this process: where the watermark overlapped with the text, the regular expression likely would have been unable to identify and remove the watermark (a false negative result), and this small difference in the watermarks (not the text) would produce different hashes.

After XML/PDF versions were retrieved, text was extracted using either the `xml2` package (59) or `pdftools`. Figure 4 and Table 2 show the number of fulltext documents included in the corpus. All together, the corpus includes 34,896 documents from 6 journals.

Table 2: Document counts, by journal, and years included in the corpus.

journal	n	start	end
Behavior Genetics	2268	1970	2010
Intelligence	1237	1977	2010
Mankind Quarterly	1821	1960	2004
Personality and Individual Differences	7274	1980	2010
Psychological Reports	21398	1960	2010

journal	n	start	end
Behavioral and Brain Sciences	898	1978	2010

Text preparation

After document retrieval and text extraction, we pre-processed the text using the `spaCy` NLP (natural language processing) Python library (60 version 1.9.0) and the R API `spacyr` (61 version 1.2.1). Specifically, we applied regular expressions to remove header/footer copyright notices and hyphenation, used spaCy to annotate and extract noun phrases (eg, “the intelligence test items”), and then cleaned and standardized these phrases (eg, removing the/an/a, converting all text to lowercase, and replacing all whitespace with underscores: “intelligence_test_items”). We then counted the occurrence of each noun phrase in the document. The aggregated “document-term matrix” was stored in Parquet format (62) for performance reasons, and written and read using the `arrow` package for R (63 version 7.0.0).

NLP-extracted noun phrases offer a number of advantages over the more traditional unigram (“single word”) terms. First, noun phrase extraction removes many standard stopwords (articles, common verbs, numbers) without relying on a fixed, *a priori* stopword list. Phrases can be more informative than single terms, for example, distinguishing “intelligence test” from “hypothesis test.” While simple n-gram extraction will include numerous phrases that are not especially meaningful. Consider the sentence “Since our first analyses of feeding patterns in rats, we had been using a criterion of 40 minutes (Le Magnen & Tallon 1963; 1966)”(64). Bigrams such as “since our” and “criterion 40” will likely be discarded in vocabulary selection, but significantly increase the computational cost of vocabulary selection. Noun phrase extraction is therefore more efficient.

After noun phrase extraction, the corpus comprised 43,162,055 total tokens of 6,320,783 distinct phrases.

Data and code availability

Document metadata retrieved from Crossref does not appear to be covered by any copyright or other intellectual property restrictions. However, due to copyright restrictions, we are unable to make document fulltext or Web of Science search results publicly available. Code for the corpus assembly and text preparation steps described above is available by request.

The public analysis repository, *[URL and DOI will go here]*, includes document metadata and documentwise counts of NLP-extracted noun phrases (“document-term matrices”), along with code to reproduce the analysis of the following sections.

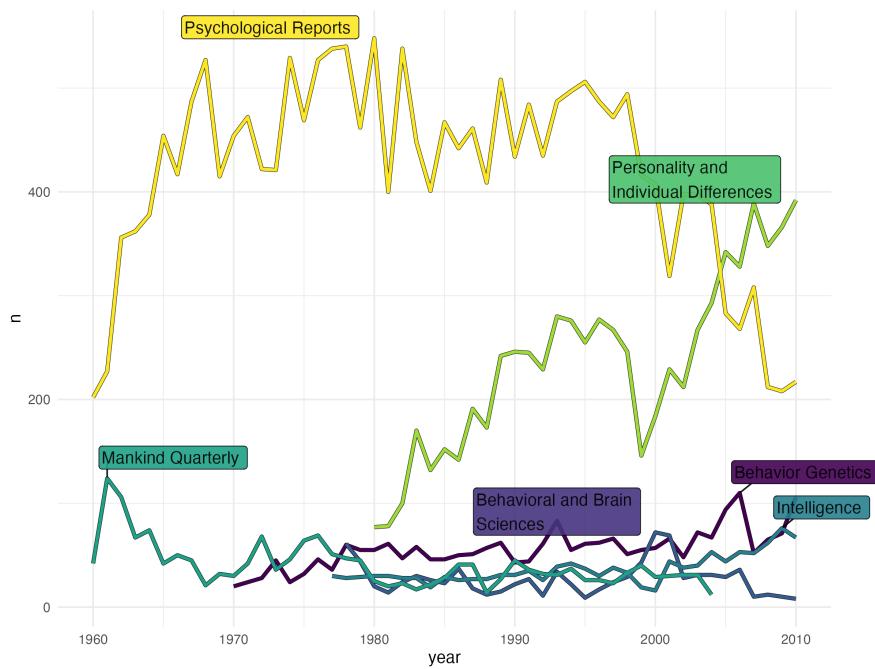


Figure 4: Count of documents in the corpus, by journal and year.

Vocabulary selection

We took an information-theoretic approach to vocabulary selection (65). Consider a game in which I draw a document from the corpus, then a single token from that document. I tell you the term, and you have to guess which document I picked. Intuitively, highly informative terms (in this project, noun phrase types) are distinctive, allowing you to dramatically narrow down the list of potential documents. This “informativeness” of a term can be quantified as the KL (Kullback-Leibler) divergence from a “baseline” uniform distribution across documents to the distribution conditional on the term, which we abbreviate as ΔH . Because the most informative terms tend to be typos and OCR errors — these are unique to a single document — we multiply the KL divergence by the logarithm of the total number of occurrences of the term across the entire corpus. We refer to the resulting measure as $\log(n)\Delta H$ or ndH .

In the current project, we found that this ndH approach heavily favored recurrent noun phrases in the longest documents. Many of the documents published in *Psychological Reports* are extremely short, 1-2 page brief notices of a single study; while many of the documents published in *Brain and Behavioral Sciences* are book-length collections that include a long review article and sometimes dozens of commentaries. Very generic noun phrases that happen to appear in BBS can occur orders of magnitude more often than highly distinctive phrases in PS, and so the $\log(n)$ factor overwhelms the ΔH factor.

To address this, we used a different baseline distribution of documents, namely, one in which document probability is proportional to length. This makes phrases from short documents much more “surprising” (much less likely to occur according to the baseline), and hence substantially increases their informativeness. This was more effective at identifying useful phrases from across the corpus.

A common rule of thumb in topic modeling is that the vocabulary should have about 10 times as many distinct terms as the number of documents in the corpus. However, we had some concerns with computational demands here: the resulting document-term matrix would have roughly $10 \times 33,000^2$ or 10.9 billion entries; with 10% density this would require on the order of 4 GB of memory just for a single copy of the matrix; and most of the analysis was to be conducted on the authors’ laptops. We therefore chose to work with three smaller vocabularies, $5 \times$, $1 \times$ and $\frac{1}{5} \times$ the number of documents. We refer to these as the “large” (174,480 distinct phrases), “medium” (34,896) and “small” (6,979) vocabulary, respectively. We compare findings across vocabularies as a robustness check.

Topic modeling

To fit topic models, we followed the approach proposed by Rohe and Zeng (66), which uses varimax-rotated partial principal components instead of the variational inference methods used by standard topic model packages such as `stm` (67). This novel approach was implemented in the R package [*redacted for anonymous review*] .

Topic models were fit for all three vocabularies (large, medium, and small) with $k = 5, 10, 20, \dots, 70$ (number of topics), resulting in a total of $24 = 3 \times 8$ models.

Following the approach of (65), we did not attempt to identify a unique best fitted model for further analysis. While the manuscript text focuses on the medium vocabulary, $k = 40$ model as the “median” among the 24 fitted models, we compare and contrast findings from this model with those from the other models.

Topic model interpretation and quality assessment

After fitting topic models, our first research question was whether we could identify distinctive “race science discourse” topics. Figures [ED-1](#), [ED-2](#), and [ED-3](#) show the gamma (topic-document) distributions for each value of k for the three vocabularies. In these figures, each panel corresponds to a single journal- k combination, each row of cells is a single document in that journal, and each column of cells is a single topic in that particular model. Topics correspond within columns of panels, but not within rows; for example, topic 04 for $k = 10$ does not necessarily correspond to topic 04 for $k = 20$. Color intensity indicates the value of gamma for that particular topic-document combination.

For $k = 5, 10$ the distributions are quite noisy, difficult to interpret, and generally don’t form very strong visual clusters. As k increases from 20 through 40, coherent bands start to appear for journal-distinctive topics in MQ, BG, Int and, to a lesser degree, PID. Note that the distinctive topics for MQ and Int do not obviously coincide. Above about $k = 50$, the models appear to identify more fine-grained topics and the coherent bands fade.

We next constructed Silge plots (18), showing the top 15 (highest-probability) terms (noun phrases) from each topic in the model; see figures [ED-4](#), [ED-5](#), [ED-6](#). Perusing these term lists, we focused on topics in each model that contained racial terms. Every fitted topic model included at least one racial topic. Importantly, in some cases these topics contained terms related to intelligence research, but in other cases they did not.

The gamma distributions and Silge plots suggested that, once k was sufficiently large, the topic models were distinguishing between two types of race science articles. We therefore examined the prevalence of topics that used racial terms, intelligence-research terms, or both, by journal and across all 24 models; see figures [ED-7](#), [ED-8](#), [ED-9](#). In these visualizations, each row of facets corresponds to a value of k (excluding 5 as too noisy), and each column of facets corresponds to one topic. Topics are clustered based on whether the top 15 terms contain racial terms (“race”), terms related to intelligence research (“intelligence”), or both. Line plots show the count of articles with $\gamma > 0.50$ for the given topic, by journal; thick lines are 5-year running averages and thin lines are raw counts.

Across all three vocabularies, for sufficiently large values of k the models identify two or more different racial topics, one that appears all but exclusively in

MQ and at least one that appears primarily in mainstream journals.

We next conducted a topic quality check of topic 24 from the medium vocabulary, $k = 40$ model, which appeared to identify race science research on intelligence published in mainstream journals. A spreadsheet of all articles that had a maximum value of γ for this topic was extracted (excluding articles published in *Psychological Reports*), and the top 121 articles ($\gamma > 0.97$) were reviewed manually by both authors. We coded each article as *race science discourse*, or not, using our definition of race science discourse as treating race science as a legitimate area of scientific research; this includes methodological critiques of race science and empirical tests that falsify race science hypotheses.

For the first round of review, both authors worked independently, dichotomously coding each article as race science discourse or not. We calculated the “false positive rate” — documents with maximum γ in this topic that were not race science discourse — and interrater reliability using Cohen’s κ . The consensus true positive rate was 89.3% (108/121), while the consensus false positive rate was 8.3% (10/121), with 98% agreement across the two raters ($\kappa = 0.86$). Coding spreadsheets are included in the supplemental materials S8-S9. For each of the 10 consensus false positive documents, their inclusion in the topic was readily explained, but in different ways for different documents. Some included text from other documents, as when an article started in the middle of a page; others discussed perceptions of race relations or racial animus; one piece in BG discussed differences in the herding behavior of two cattle breeds, referred to as “white” and “black.”

The 3 non-consensus documents reflected essential ambiguity in the operationalization of race science discourse. One is a table of contents from a 1980 issue of Int; arguably this document should have been excluded as irrelevant front matter, but arguably it legitimizes the research published in the issue. The second case is a paper on racial differences in emotional intelligence that was classified as race science discourse by one author, while the other author felt it was more focused on psychometrics (68). And a 1990 article in MQ (<https://www.unz.com/print/MankindQuarterly-1990q3-00108/>) proposed to offer an economic explanation for crime. While it has a few pages connecting race/ethnicity to both crime and poverty, it ultimately focuses on unemployment and alcohol sales, rather than race, as the key predictors of crime.

Based on this qualitative review, we judged that the topic model approach was sufficiently sensitive (low “false positive” rate) in identifying race science discourse. However, “false negatives” are still possible, as some documents that human readers would classify as race science discourse might have been associated with other topics.

Supplemental analysis of Behavior Genetics

To check the finding that very few race-and-intelligence documents had been published in BG, we ran an independent search for documents with the keywords

“race” and “intelligence” published in BG from 1972 to 2020 using Springer’s journal search website. This search returned 125 documents. Thirty-three were presentation abstracts from meetings of the Behavior Genetics Association; we did not examine these further. Fourteen of the remaining documents had been published since 2010 (after the scope of the primary study). Among these, one obituary and two papers celebrated the work of John Loehlin, a prominent race-and-intelligence researcher who had been Director of the American Eugenics Society from 1968-1972 (69–71). Another was a retrospective of the work of Lindon Eaves, a geneticist who made at least one notable contribution to debates on the Scarr-Rowe hypothesis (interactions between heritability, race, and class). None of these 14 documents reported any studies of racial differences.

Twenty-eight documents in this sample from BG were published between 1990 and 2010. One was an obituary of Jerry Hirsch, a critic of hereditarianism in general and Jensen in particular (72), and another was an obituary of David Rowe (73). Only 2 of these 28 documents reported race differences of any kind: one examining interactions among race, sex, and heritability for adolescent BMI (Body Mass Index, used as a measure of overweight/obesity) (74); and the other interactions among race, family history of alcoholism, and visuospatial performance (75). Finally, Philip Vernon — one of the Pioneer-funded scientists we identified earlier — published a critical review of a book that attempted to address hereditarianism, the Flynn effect, race-and-intelligence research, and some related issues.

At the same time, we were unable to find any articles published in BG that acknowledged the involvement of the field with scientific racism with anything like the force of the statements that have been made by biological anthropologists (33) and human geneticists (53). Notably, the ASHG report includes discussions of behavior genetics race science, and in particular is critical of the field and the organization for failing to publicly reject the race-and-intelligence claims made by Shockley and Jensen (53). A search for “eugenics” in BG found articles that presented the issue as part of the distant past, and a search for “racism” turned up studies of racist attitudes. A statement on the Behavior Genetics Association website, dated 2021, acknowledges that “The history of our field is inextricably linked with racism, including the misuse of behavior genetic research to support violent eugenic policies,” but primarily focuses on BGA member demographics (https://www.bga.org/content.aspx?page_id=22&club_id=971921&module_id=567723).

Limitations

A key limitation of the current study stems from the inability of text mining techniques to draw on a broader cultural context than what is represented in the corpus. Scholars of racism and US race relations have noted a shift in the way racist attitudes are understood and expressed, from overt and direct to a subtle “color-blind” racism (76). Color-blind racism represents a change in the primary mechanisms by which white privilege is maintained, using non-racialized language in furtherance of the racial status quo both institutionally

and individually, and increasingly reliant on the invisibility of socio-cultural mechanisms to preserve inequalities. Color-blind racism involves the use of racialized “codes” or “dogwhistles,” language that is superficially non-racial but carries racial implications, such as (in the US context) “violent inner city criminals” or “welfare queen” (77). Alderfer notes that the racist implications of *The Bell Curve* only emerge gradually, with the opening of the book using the color-blind language of “group differences” (78). These cases suggest further that color-blindness can facilitate white supremacist appropriation of non-racial research — even contrary to the researchers’ own intentions — as scientific racist readers re-interpret racial-neutral language as dogwhistles (26,79).

Techniques such as topic modeling may be able to account for discursive shifts from explicit to coded racial language, but only if the corpus contains a critical mass of documents using both sets of terms together. This would seem to require either a relatively gradual transition, the inclusion of reflective commentary — such as scholarship identifying and analyzing the transition — or both. It is unlikely that these requirements are met by the corpus used in this study. So it is plausible that the topic models used here have some rate of “false negatives,” documents in the corpus that engage in race science discourse but in a color-blind way that avoids the use of the explicit racial terms in the race-and-intelligence topic.

The phenomenon of color-blind racism is closely related to ambiguities that cannot be resolved even by human coders familiar with the broader cultural context. Two relevant examples are the “Flynn effect,” a secular increase in intelligence test scores noted by philosopher James Flynn; and studies of “national IQ,” exemplified by the work of Richard Lynn. The Flynn effect implies that environmental factors can create group differences in IQ that are comparable to Black-White differences, and thus directly challenges claims that the racial differences are biological. Using methods widely regarded as ad hoc, cherry-picking, and generally unreliable, Lynn and collaborators have claimed that national IQ averages are correlated with national GDP. Both the Flynn effect and “national IQ” are strongly associated with race-and-intelligence discourse in the US cultural context. But scientific journal articles on these topics may not include any racial language at all. For example, a meta-analysis of the Flynn effect does not discuss racial differences or use terms such as “White” or “Black” outside the reference list (80). Should such articles be coded as race science discourse? We expect that even readers who accept our definitions of race science and race science discourse will reasonably disagree on how to answer this question. Topic models, of course, are not even capable of representing this essential ambiguity.

References

1. Barkan, E. *The Retreat of Scientific Racism: Changing Concepts of Race in Britain and the United States Between the World Wars*. (Cambridge University Press, 1992).

2. Duster, T. *Backdoor to Eugenics*. (Routledge, 2003).
3. Panofsky, A., Dasgupta, K. & Iturriaga, N. How White nationalists mobilize genetics: From genetic ancestry and human biodiversity to counterscience and metapolitics. *American Journal of Physical Anthropology* **175**, 387–398 (2021).
4. Pronczuk, M. & Ryckewaert, K. A Racist Researcher, Exposed by a Mass Shooting. *The New York Times* (2022).
5. Jackson Jr., J. P. Arthur Jensen, evolutionary biology, and racism. *History of Psychology* No Pagination Specified–No Pagination Specified (2022). doi:[10.1037/hop0000221](https://doi.org/10.1037/hop0000221)
6. Mehler, B. Foundation for fascism: The new eugenics movement in the United States. *Patterns of Prejudice* **23**, 17–25 (1989).
7. Winston, A. S. Science in the Service of the Far Right: Henry E. Garrett, the IAAEE, and the Liberty Lobby. *Journal of Social Issues* **54**, 179–210 (1998).
8. Tucker, W. H. *The funding of scientific racism: Wickliffe Draper and the Pioneer Fund*. (University of Illinois Press, 2002).
9. Schaffer, G. ““Scientific” Racism Again?": Reginald Gates, the Mankind Quarterly and the Question of ‘Race’ in Science after the Second World War. *Journal of American Studies* **41**, 253–278 (2007).
10. Saini, A. *Superior: The Return of Race Science*. (Beacon Press, 2019).
11. Winston, A. S. in *Oxford Research Encyclopedia of Psychology* (Oxford University Press, 2020). doi:[10.1093/acrefore/9780190236557.013.516](https://doi.org/10.1093/acrefore/9780190236557.013.516)
12. Adams, D. & Pilloud, M. The (Mis)appropriation of Biological Anthropology in Race Science and the Implications for Forensic Anthropology. *FA* (2021). doi:[10.5744/fa.2021.0010](https://doi.org/10.5744/fa.2021.0010)
13. Saini, A. Draper’s Millions: The Philanthropic Wellspring of Modern Race Science. (2022).
14. Jackson, J. P., Jr. *Science for Segregation: Race, Law, and the Case Against Brown V. Board of Education*. (NYU Press, 2005).
15. Jackson, J. P., Jr. & Winston, A. S. The Mythical Taboo on Race and Intelligence. *Review of General Psychology* (2020). doi:[10.1177/1089268020953622](https://doi.org/10.1177/1089268020953622)
16. Miller, A. The Pioneer Fund: Bankrolling the Professors of Hate. *The Journal of Blacks in Higher Education* 58–61 (1994). doi:[10.2307/2962466](https://doi.org/10.2307/2962466)
17. Beirich, H. Pioneer Fund Assets Divided; New Leadership Appointed. *Southern Poverty Law Center* (2013).
18. in *Text Mining with R: A Tidy Approach* ch. 6 ("O'Reilly Media, Inc.", 2017).

19. Mehler, B. Beyondism: Raymond B. Cattell and the new eugenics. *Genetica* **99**, 153–163 (1997).
20. Jensen, A. How Much Can We Boost IQ and Scholastic Achievement? *Harvard Educational Review* **39**, 1–123 (1969).
21. Panofsky, A. *Misbehaving Science: Controversy and the Development of Behavior Genetics*. (University of Chicago Press, 2014).
22. Scarr, S. Three cheers for behavior genetics: Winning the war and losing our identity. *Behav Genet* **17**, 219–228 (1987).
23. Slobodian, Q. The Unequal Mind: How Charles Murray and Neoliberal Think Tanks Revived IQ. *Capitalism: A Journal of History and Economics* **4**, 73–108 (2023).
24. Schroeder, S. A. The Limits of Democratizing Science: When Scientists Should Ignore the Public. *Philosophy of Science* **89**, 1034–1043 (2022).
25. Tabery, J. Why Is Studying the Genetics of Intelligence So Controversial? *Hastings Center Report* **45**, S9–S14 (2015).
26. Gillborn, D. Softly, softly: Genetics, intelligence and the hidden racism of the new geneism. *Journal of Education Policy* **31**, 365–388 (2016).
27. Carlson, J. & Harris, K. Quantifying and contextualizing the impact of bioRxiv preprints through automated social media audience segmentation. *PLOS Biology* **18**, e3000860 (2020).
28. Henn, B. M., Merchant, E. K., O'Connor, A. & Rulli, T. Why DNA Is No Key to Social Equality: On Kathryn Paige Harden's 'The Genetic Lottery'. *Los Angeles Review of Books* (2021).
29. Anthropology, O. AAPA statement on biological aspects of race. *American Journal of Physical Anthropology* **101**, 569–70 (1996).
30. Human Genetics, A. S. of. ASHG denounces attempts to link genetics and racial supremacy. *American Journal of Human Genetics* **103**, 636 (2018).
31. Board, A. E. AAA statement on race. *American Anthropologist* **100**, 712–713 (1998).
32. Social Anthropologists of the UK, A. of. Statement on racism by the committee of the ASA. Retrieved from (2020).
33. Fuentes, A. et al. AAPA Statement on Race and Racism. *American Journal of Physical Anthropology* **169**, 400–402 (2019).
34. Institute, N. H. G. R. Race. Retrieved September 5, 2023 from (2023).
35. Rotimi, C. ASHG statement regarding the warping of genetic knowledge to feed racist ideology. Retrieved from (2022).

36. Wynshaw-Boris, A. American society of human genetics statement regarding concepts of ‘good genes’ and human genetics. *Retrieved from* (2020).
37. Association, A. P. Apology to people of color for APA’s role in promoting, perpetuating, and failing to challenge racism, racial discrimination, and human hierarchy in US. *Retrieved from* 10, 2021 (2021).
38. Baron, J. RE: OMB-2023-0001, ”initial proposals for updating OMB’s race and ethnicity statistical standards”. *Retrieved from* (2023).
39. [Science must respect the dignity and rights of all humans.](#) *Nat Hum Behav* 6, 1029–1031 (2022).
40. Psychological Science, A. for. Statement on racism, bias, and intolerance. *Retrieved from* (2021).
41. Society, P. Statement on racism. *Retrieved from* (2020).
42. Jackson Jr, J. P. & Winston, A. S. The mythical taboo on race and intelligence. *Review of General Psychology* 25, 3–26 (2021).
43. Roberts, S. Dealing with diversity in psychology: Science and ideology. (2022). doi:10.31234/osf.io/xk4yu
44. Herbert, J. M. Academic free speech or right-wing grievance? *Digital Discovery* 2, 260–297 (2023).
45. Meyer, M. N. *et al.* Wrestling with Social and Behavioral Genomics: Risks, Potential Benefits, and Ethical Responsibility. *Hastings Center Report* 53, S2–S49 (2023).
46. Block, N. J. & Dworkin, G. [IQ, Heritability and Inequality, Part 2.](#) *Philosophy & Public Affairs* 4, 40–99 (1974).
47. Douglas, H. E. *Science, policy, and the value-free ideal.* (University of Pittsburgh Press, 2009).
48. Kitcher, P. [An Argument About Free Inquiry.](#) *Noûs* 31, 279–306 (1997).
49. Kourany, J. A. [Should Some Knowledge Be Forbidden? The Case of Cognitive Differences Research.](#) *Philosophy of Science* 83, 779–790 (2016).
50. Ogbunugafor, C. B. DNA, basketball, and birthday luck. A review of The genetic lottery: Why DNA matters for social equality. By Kathryn Paige Harden, 2021, US: Princeton University Press. 312 pp. ISBN: 9780691190808 (Hardcover). *American Journal of Biological Anthropology* 179, 501–504 (2022).
51. Wenzel, M., Okimoto, T. G., Feather, N. T. & Platow, M. J. [Retributive and Restorative Justice.](#) *Law Hum Behav* 32, 375–389 (2008).
52. Wormer, K. van. in *Encyclopedia of Social Work* (2013). doi:10.1093/acrefore/9780199975839.013.345

53. Jackson, C. S., Turner, D., June, M. & Miller, M. V. [Facing Our History—Building an Equitable Future](#). *The American Journal of Human Genetics* **110**, 377–395 (2023).
54. Ooms, J. [Pdftools: Text Extraction, Rendering and Converting of PDF Documents](#). (2023).
55. Chamberlain, S., Zhu, H., Jahn, N., Boettiger, C. & Ram, K. [Rcrossref: Client for Various 'CrossRef' APIs](#). (2019).
56. Flynn, J. R. [Evidence against Rushton: The genetic loading of WISC-R subtests and the causes of between-group IQ differences](#). *Personality and Individual Differences* **26**, 373–379 (1998).
57. Rushton, J. P. [Secular gains in IQ not related to the g factor and inbreeding depression – unlike Black-White differences: A reply to Flynn](#). *Personality and Individual Differences* **26**, 381–389 (1998).
58. Flynn, J. R. [Reply to Rushton: A gang of gs overpowers factor analysis](#). *Personality and Individual Differences* **26**, 391–393 (1998).
59. Wickham, H., Hester, J., Ooms, J., RStudio & example), R. F. (Copy of R. homepage cached as. Xml2: Parse XML. (2023).
60. spaCy: Industrial-strength Natural Language Processing (NLP) with Python and Cython. (2018).
61. Benoit, K., Matsuo, A. & Council (ERC-2011-StG 283794-QUANTESS), E. R. Spacyr: Wrapper to the 'spaCy' 'NLP' Library. (2020).
62. Vohra, D. in *Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools* (ed. Vohra, D.) 325–335 (Apress, 2016). doi:[10.1007/978-1-4842-2199-0_8](https://doi.org/10.1007/978-1-4842-2199-0_8)
63. Richardson, N. *et al.* Arrow: Integration to 'Apache' 'Arrow'. (2023).
64. Sclafani, A. [Correlation and causation in the study of feeding behavior](#). *Behavioral and Brain Sciences* **4**, 590–591 (1981).
65. Hicks, D. J. [Productivity and interdisciplinary impacts of Organized Research Units](#). *Quantitative Science Studies* **2**, 990–1022 (2021).
66. Rohe, K. & Zeng, M. Vintage Factor Analysis with Varimax Performs Statistical Inference. (2020). doi:[10.48550/arXiv.2004.05387](https://doi.org/10.48550/arXiv.2004.05387)
67. Roberts, M. E., Stewart, B. M. & Tingley, D. [Stm: An R Package for Structural Topic Models](#). *Journal of Statistical Software* **91**, 1–40 (2019).
68. Gignac, G. E. & Ekermans, G. [Group differences in EI within a sample of black and white South Africans](#). *Personality and Individual Differences* **49**, 639–644 (2010).
69. [John C. Loehlin \(1926–2020\)](#). *Behav Genet* **50**, 385–386 (2020).
70. Waldman, I. & Turkheimer, E. [Introduction to a Festschrift for John Loehlin](#). *Behav Genet* **44**, 547–548 (2014).

71. Plomin, R. **Genotype-Environment Correlation in the Era of DNA**. *Behav Genet* **44**, 629–638 (2014).
72. Roubertoux, P. L. **Jerry Hirsch (20 September 1922–3 May 2008): A Tribute**. *Behav Genet* **38**, 561–564 (2008).
73. Rodgers, J. L., Jacobson, K. & van den Oord, E. **Obituary: David Christian Rowe**. *Behav Genet* **33**, 627–628 (2003).
74. Jacobson, K. C. & Rowe, D. C. **Genetic and Shared Environmental Influences on Adolescent BMI: Interactions with Race and Sex**. *Behav Genet* **28**, 265–278 (1998).
75. Berman, S. M. & Noble, E. P. **Reduced visuospatial performance in children with the D2 dopamine receptor A1 allele**. *Behav Genet* **25**, 45–58 (1995).
76. Bonilla-Silva, E. *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States*. (Rowman & Littlefield Publishers, 2006).
77. Saul, J. in *New Work on Speech Acts* (eds. Fogal, D., Harris, D. W. & Moss, M.) 0 (Oxford University Press, 2018). doi:[10.1093/oso/9780198738831.003.0013](https://doi.org/10.1093/oso/9780198738831.003.0013)
78. Alderfer, C. P. **The science and nonscience of psychologists' responses to The Bell Curve**. *Professional Psychology: Research and Practice* **34**, 287–293 (2003).
79. Wills, M. **Are Clusters Races? A Discussion of the Rhetorical Appropriation of Rosenberg et al.'s 'Genetic Structure of Human Populations'**. *Philosophy & Theory in Biology* **9**, (2017).
80. Trahan, L., Stuebing, K. K., Hiscock, M. K. & Fletcher, J. M. **The Flynn Effect: A Meta-analysis**. *Psychol Bull* **140**, 1332–1360 (2014).

Extended Data

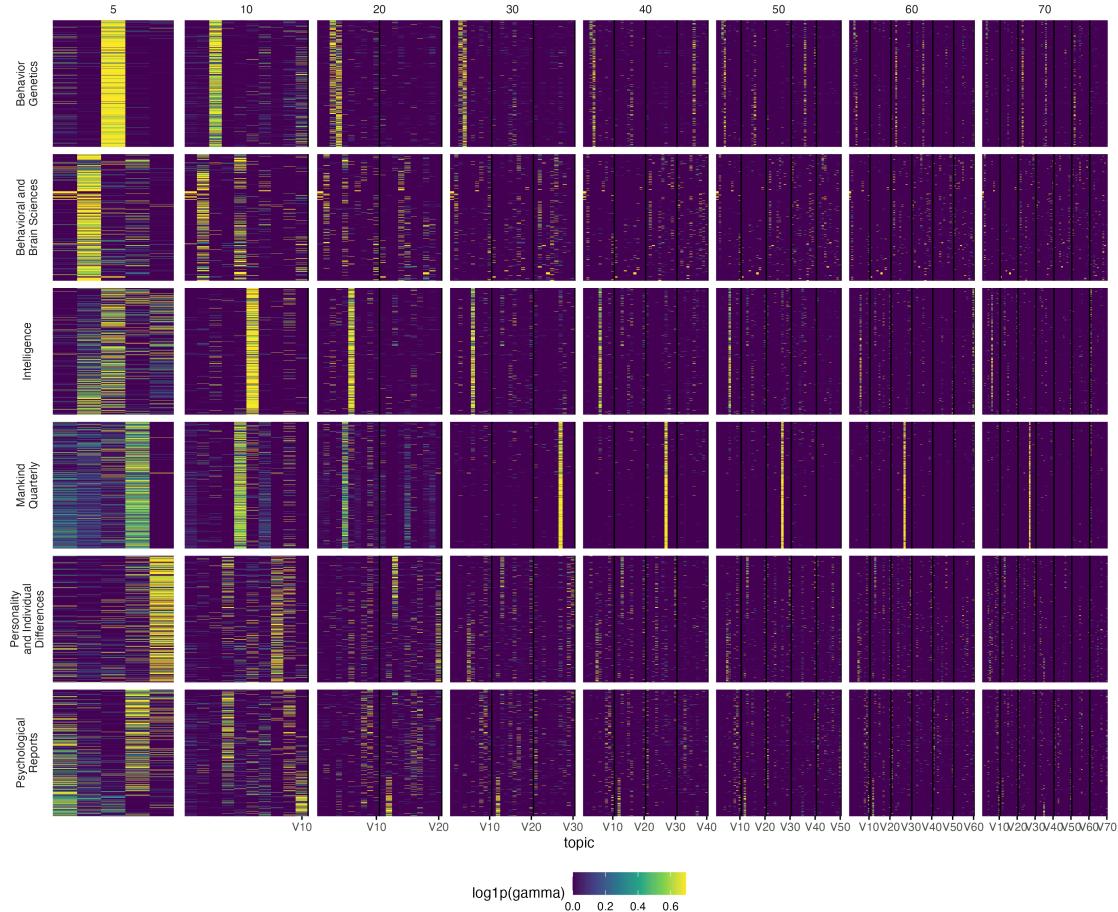


Figure ED-1: Fitted topic models: Gamma (topic-document) distributions, large vocabulary.

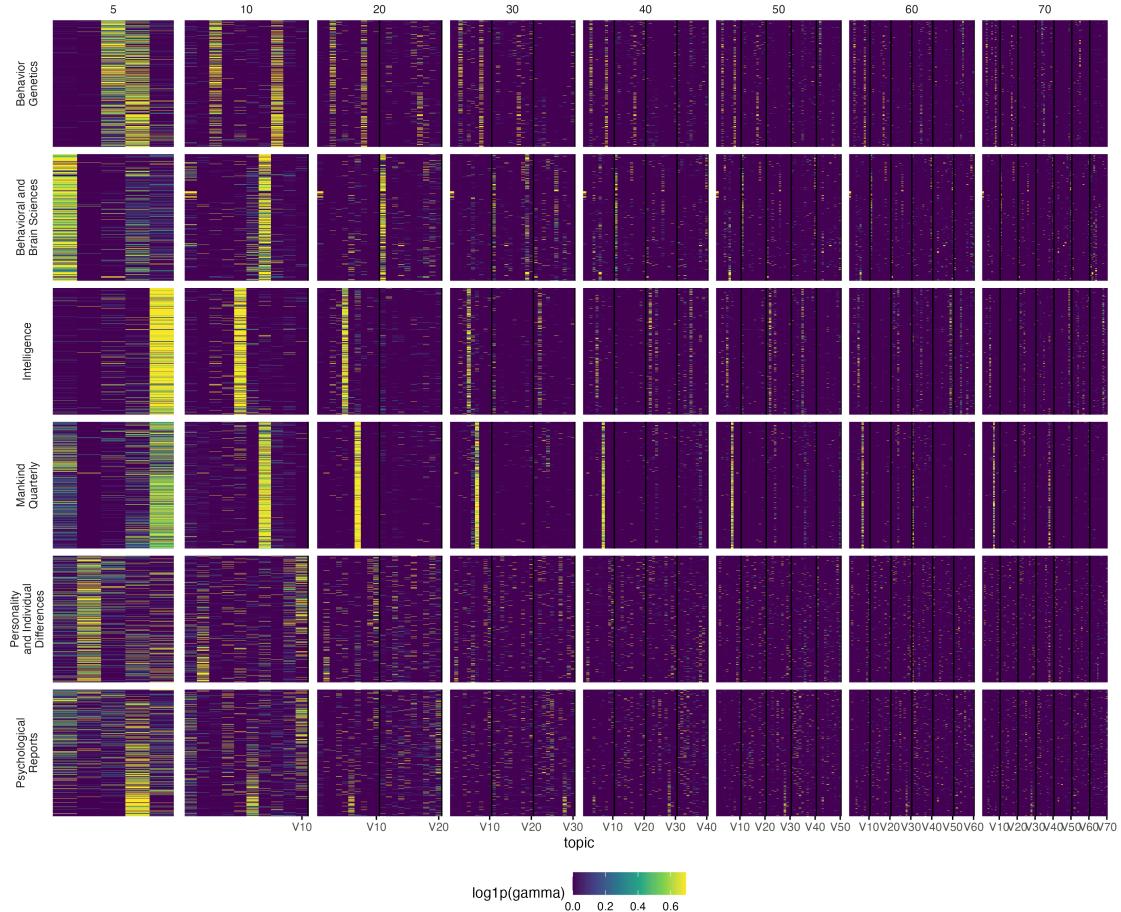


Figure ED-2: Fitted topic models: Gamma (topic-document) distributions, medium vocabulary.

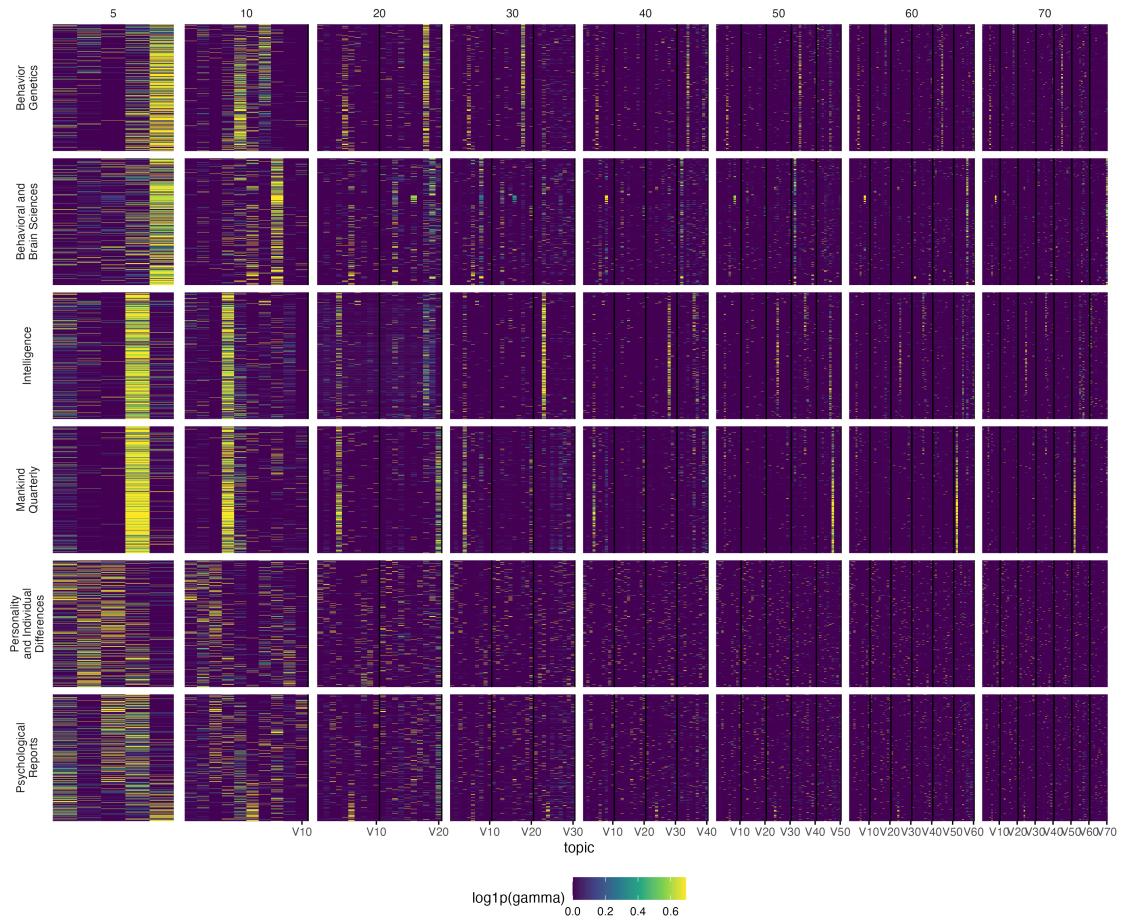


Figure ED-3: Fitted topic models: Gamma (topic-document) distributions, small vocabulary.

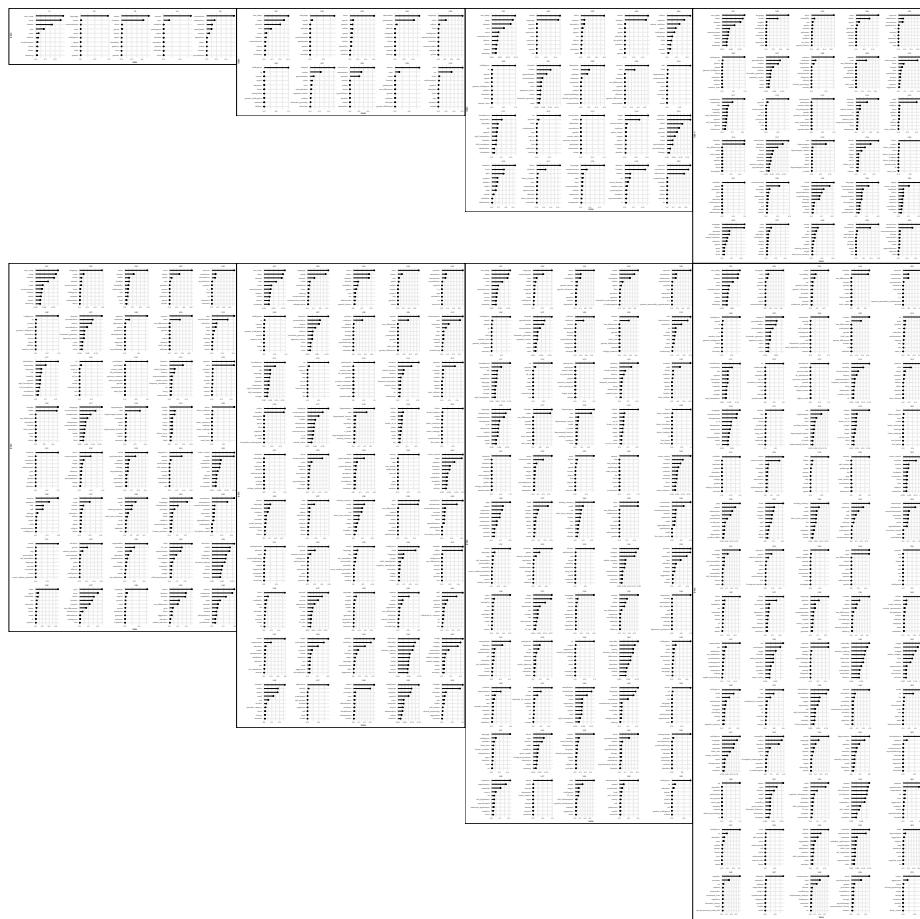


Figure ED-4: Silge plots (top 15 terms) for each topic and model, large vocabulary.

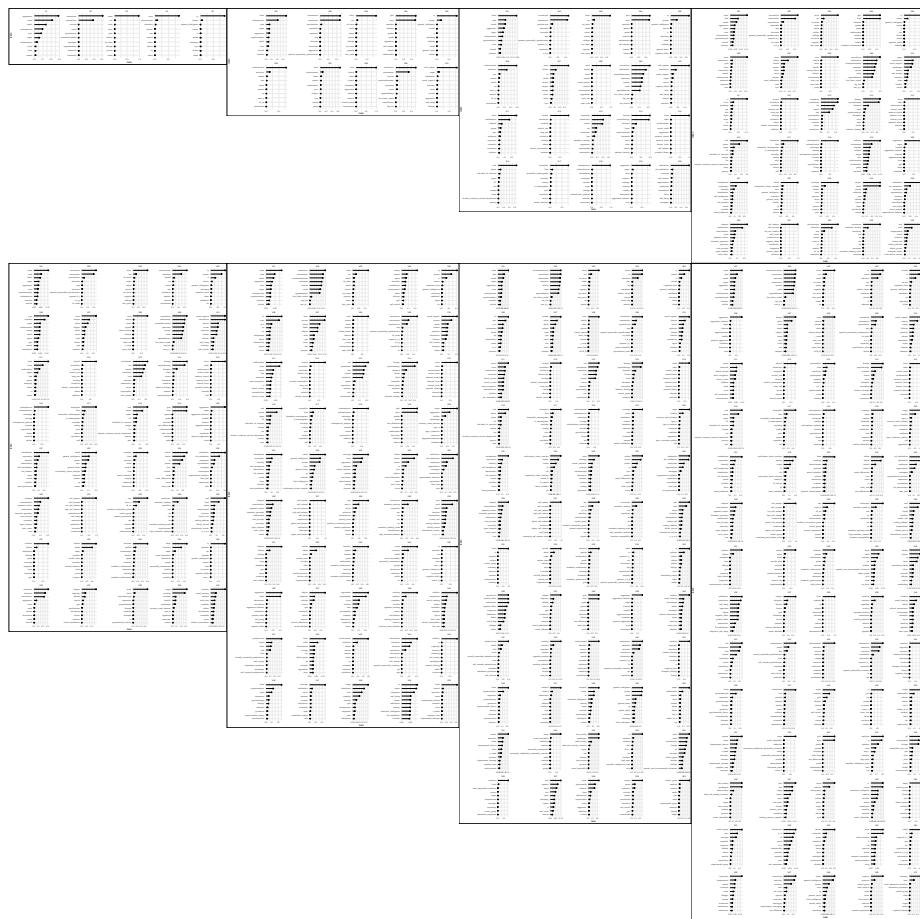


Figure ED-5: Silge plots (top 15 terms) for each topic and model, medium vocabulary.

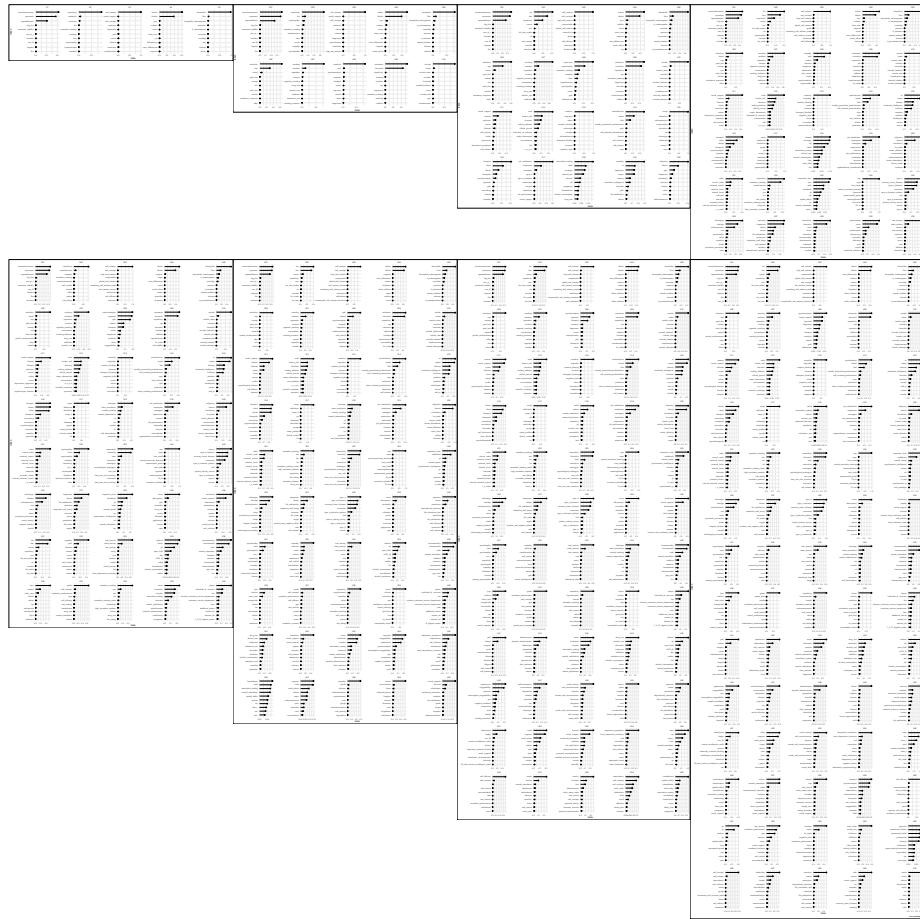


Figure ED-6: Silge plots (top 15 terms) for each topic and model, small vocabulary.

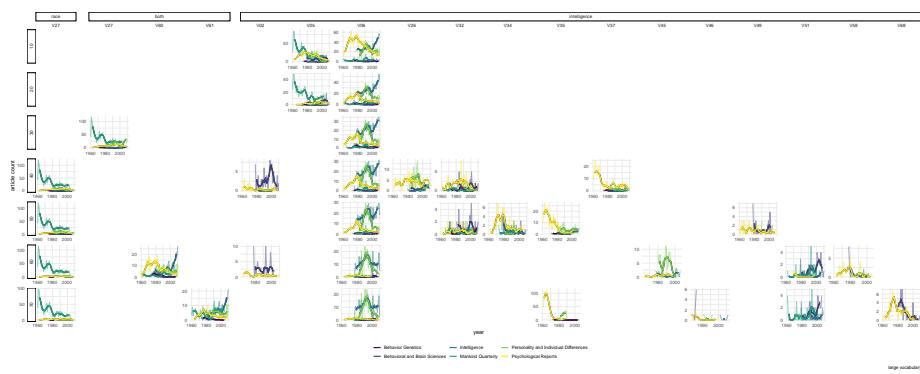


Figure ED-7: Topics with “race,” “intelligence,” or both occurring in the top 15 terms, large vocabulary.

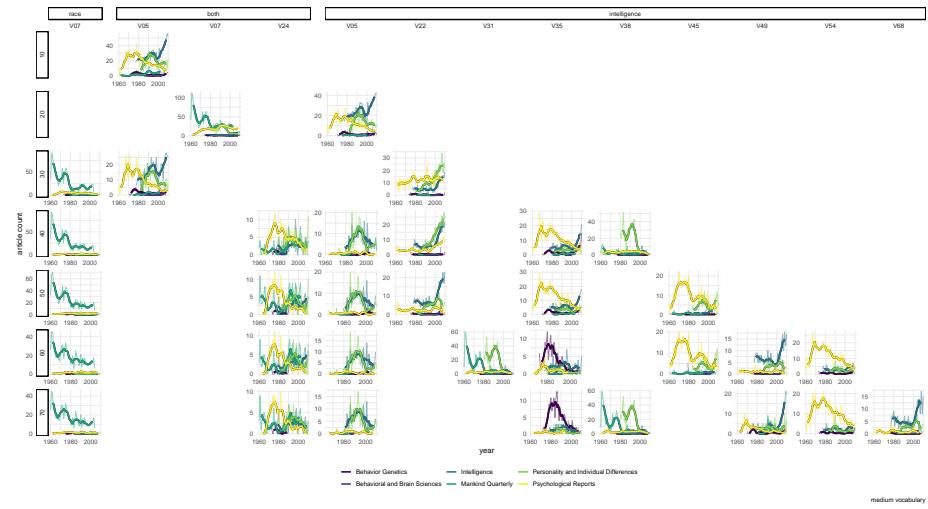


Figure ED-8: Topics with “race,” “intelligence,” or both occurring in the top 15 terms, medium vocabulary.

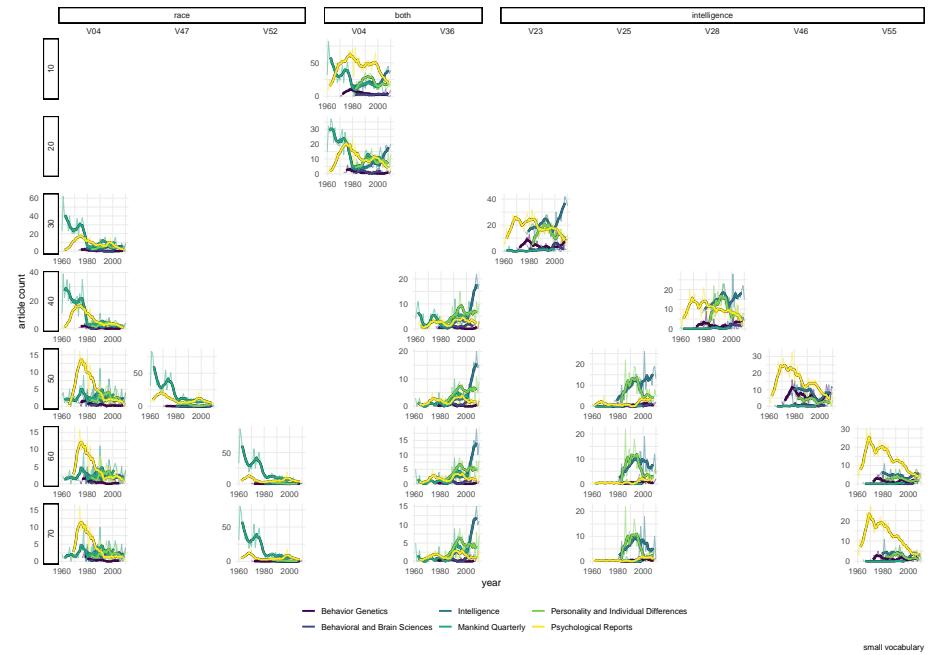


Figure ED-9: Topics with “race,” “intelligence,” or both occurring in the top 15 terms, small vocabulary.