

Research



Cite this article: Carlson J, Harris K. 2022 The apportionment of citations: a scientometric analysis of Lewontin 1972. *Phil. Trans. R. Soc. B* **377**: 20200409.

<https://doi.org/10.1098/rstb.2020.0409>

Received: 26 July 2021

Accepted: 22 December 2021

One contribution of 15 to a theme issue 'Celebrating 50 years since Lewontin's apportionment of human diversity'.

Subject Areas:

genetics, genomics, evolution, bioinformatics

Keywords:

bibliometrics, citations, Twitter, Lewontin's fallacy, population genetics, human diversity

Author for correspondence:

Kelley Harris

e-mail: harris.kelley@gmail.com

The apportionment of citations: a scientometric analysis of Lewontin 1972

Jedidiah Carlson¹ and Kelley Harris^{1,2}

¹Department of Genome Sciences, University of Washington, Seattle, WA, USA

²Computational Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

JC, 0000-0002-1363-872X; KH, 0000-0003-0302-2523

'The apportionment of human diversity' (1972) is the most highly cited research article published by geneticist Richard Lewontin in his career. This study's primary result—that most genetic diversity in humans can be accounted for by within-population differences, not between-population differences—along with Lewontin's outspoken, politically charged interpretations thereof, has become foundational to the scientific and cultural discourse pertaining to human genetic variation. The article has an unusual bibliometric trajectory in that it is much more salient in the bibliographic record today compared to the first 20 years after its publication. Here, we highlight four factors that may have played a role in shaping the paper's fame: (i) citations in influential publications across several disciplines; (ii) Lewontin's own popular books and media appearances; (iii) the renaissance of population genetics research of the early 1990s; and (iv) the serendipitous collision of scientific progress, influential books and papers, and heated controversies around the year 1994. We conclude with an analysis of Twitter data to characterize the communities and conversations that continue to keep this study at the centre of discussions about race and genetics, prompting new challenges for scientists who have inherited Lewontin's legacy.

This article is part of the theme issue 'Celebrating 50 years since Lewontin's apportionment of human diversity'.

1. Introduction

Richard Lewontin's 1972 study, 'The apportionment of human diversity' [1] (hereafter referred to as 'Lewontin 1972'), published as a contributed chapter in volume 6 of the book series, *Evolutionary biology* [2], is widely considered to be a landmark publication in human population genetics research [3–6]. This paper is widely attributed as the originator of an aphoristic sound bite that is foundational to our understanding of human genetic diversity—'there is more genetic variation within populations than between populations'. This result has been colloquially cited far beyond the measurable scope of citations in the academic literature, ranging from educational materials developed by the National Human Genome Research Institute [7] to public television documentaries [8] to numerous references on contemporary social media platforms. Furthermore, Lewontin's blunt interpretation of his results ('Human racial classification is of no social value and is positively destructive of social and human relations. Since such racial classification is now seen to be of virtually no genetic or taxonomic significance either, no justification can be offered for its continuance' [1, p. 397]) has come to be ubiquitously associated (and in some cases, incorrectly credited) with the widely held consensus that race is a social construct, not a biological one [9]. The far-reaching influence of Lewontin 1972 is partially evident in its bibliometrics: over the last half-century, it has amassed thousands of literature citations across a diverse range of disciplines; not only within Lewontin's home fields of genetics and evolutionary biology, but also anthropology, medicine [10], psychology [11], sociology [12] and information science [13], many of which have gone on to become highly influential studies within their respective fields.

Though care must be taken to avoid conflating scientometric indicators with more subjective definitions of a publication's impact and influence [14–16], a careful examination of the scientometric data surrounding Lewontin's paper is useful in helping us formulate several pertinent questions about the paper's history: what factors drove the extensive academic and cultural attention surrounding the paper, and how has that attention evolved over the last 50 years? How did the paper come to be associated with the 'more genetic variation within populations than between populations' sound bite? Even though Lewontin's contemporaries published several topically similar papers in the same era, why did Lewontin 1972 emerge as the most iconic? How has social media perpetuated and mutated the discourse surrounding this paper? Answering these questions may, in turn, spark conversations about research impact and scientific communication relevant to today's scholars: what makes an ordinary research paper have an extraordinary bibliographic and/or cultural impact? How can scientists engage with socially sensitive research topics while maintaining their personal moral and ethical convictions? What are the benefits and landmines of using social media to communicate the results and implications of such research?

In the spirit of Lewontin himself, who constantly urged his colleagues to acknowledge the interpenetration between scientific research and the socio-ecological systems in which that research is embedded [17], we attempt to address these questions by examining patterns in the citing literature, major events surrounding the broader field of human genetics research, and Lewontin's own career trajectory. We identify four factors that appear to have propelled Lewontin 1972 to its current iconic status: (i) citations in several highly influential books and papers, beginning almost immediately after its publication; (ii) Lewontin's influence through his popular books and media appearances in which he reiterated the results of his paper; (iii) rapid technical advancements in molecular genetics in the early 1990s that would prompt and enable a new generation of population geneticists to revisit landmark studies from the 1960s and 1970s; and (iv) several influential and/or provocative publications and events that coincided around the year 1994.

To better understand the ongoing cultural impacts of Lewontin 1972, we proceed to explore how the publication was referenced on Twitter during a nine-month period from 2020 to 2021 (coincidentally ending our data collection just a month before Lewontin passed away in July 2021). We found that direct references to the paper on social media are nearly non-existent, undermining the utility of standard altmetric indicators like the Altmetric Attention Score. However, an expansive corpus of tweets indirectly referencing Lewontin 1972 (which we term 'dark citations', following [18]) reveals that the concepts Lewontin presented 50 years ago continue to maintain a foothold in the cultural zeitgeist. We conclude with a discussion of how these colloquial online citations differ from the traditional bibliometric record, and how social media has, for better or worse, democratized conversations about population genetics research.

2. The bibliometric trajectory of Lewontin 1972

As of 15 June 2021, Lewontin 1972 had received 3076 citations according to Google Scholar and 1991 citations according to Semantic Scholar, translating to an average rate of 39.8–61.5

citations per year. According to Google Scholar, Lewontin's only published works with more citations are the opinion piece 'The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme', coauthored with Gould [19], and two popular science books: *The genetic basis of evolutionary change* [20] and *Not in our genes: biology, ideology, and human nature* [21]. The citation trajectory of Lewontin 1972 over time appears to be highly unusual, with only 15% of citations occurring in the first 30 years of the paper's lifespan and the remaining 85% of citations occurring since 2002 (figure 1a). The distribution of the number of citations per year is roughly bimodal, with an initial weak pulse of citations peaking in the early 1980s and tapering off to nearly 0 citations per year by the end of that decade, a pattern that was first observed by Ruvolo & Seielstad [3]. A second, much stronger, pulse of citations emerged in the early 1990s, jumped dramatically in the early 2000s, and grew steadily until the annual citation rate peaked around 2010–2015 (figure 1b).

Ruvolo & Seielstad [3] proposed two hypotheses for why Lewontin's article received so few citations early on: either (i) the scientific community had 'already come to believe that human races were effectively a scientific nonissue', rendering the results of Lewontin 1972 'obvious' or (ii) the scientific study of race and genetics was 'too politically charged' for further investigation. Ruvolo and Seielstad effectively reject the former of their hypotheses in the very first paragraph of their paper, where they state:

[Lewontin's] findings surprised those who read the paper. Although typological notions of race had been on the decline in anthropology, many scientists and laypeople continued (and a few still continue) to expect substantial genetic differences between the groups they seemed able to recognize visually [3, p. 141].

Their second hypothesis, that the study of race and genetics was considered too politically taboo in the 1970s–1980s, is also easy to reject, in light of the numerous population genetics studies during that era that probed the topic of human genetic diversity within and between races [22–32].

Although many population geneticists saw fit to investigate genetic diversity and its degree of correlation with racial classifications, they focused their studies around different technical questions and thus reached conclusions that came to be framed for lay audiences as contradictory with one another. On one hand, Lewontin, Latter, Nei and Roychoudhury focused on single genetic loci and found that the typical locus is rarely able to differentiate phenotypically defined human races. Lewontin prominently interpreted this as evidence that phenotypically defined human races are not 'taxonomically significant'. By contrast, Smouse, Spielman and Mitton studied larger sets of typical genetic markers and found that these *are* generally well powered to classify individuals into populations of origin, despite reflecting a small fraction of the overall genetic variation. As noted by Rosenberg [5] and Novembre [6], these multi-locus results do not contradict Lewontin's at a technical level, but were used to argue for the seemingly contradictory conclusion that phenotypically defined human races *are* taxonomically significant to a degree. To complicate matters further, Nei & Roychoudhury [32] repeatedly confirmed Lewontin's finding of more variation within populations than between populations, but also maintained that the

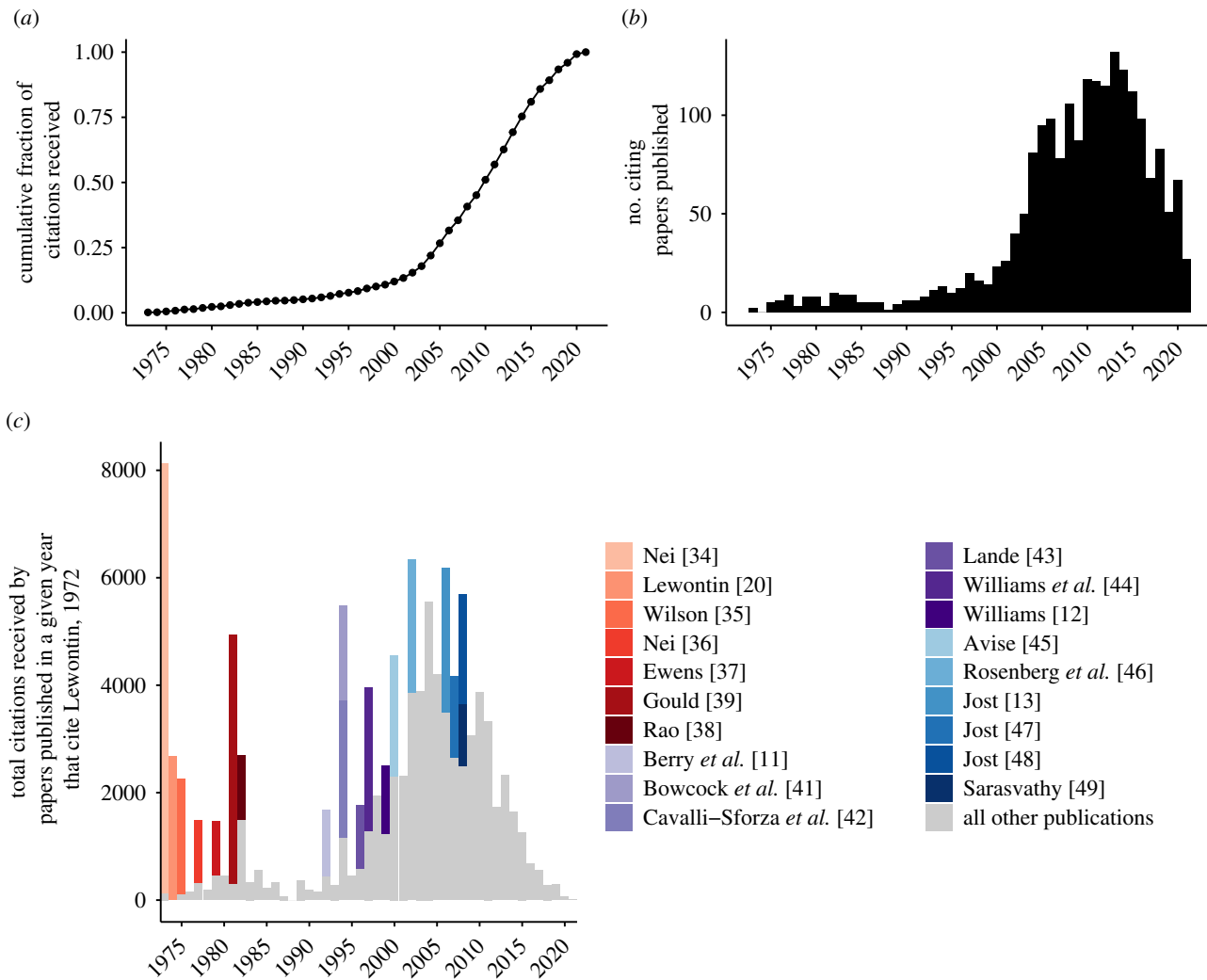


Figure 1. Bibliometric summary of Lewontin 1972. (a) Cumulative distribution of citations over time; (b) histogram of citations per year; (c) histogram of 2nd-degree citations (i.e. among citing articles in a given year, the total number of citations they have received to date) over time. Contributions from the most highly influential papers (articles that went on to receive greater than 1000 citations, according to Semantic Scholar) that cited Lewontin 1972 are indicated in coloured bar segments.

non-zero between-population variance justified ‘the existence of a biological basis for the classification of human races’.

The bibliographic manifestation of this debate appears to terminate rather abruptly in the early 1980s with a stalemate. Two papers, Nei & Roychoudhury’s [32] ‘Genetic relationship and evolution of human races’ and Ryman *et al.*’s [33] ‘Differences in the relative distribution of human gene diversity between electrophoretic and red and white cell antigen loci’ confirmed the earlier variance partitioning results, whereas Smouse *et al.* [31] maintained that a multi-locus approach could be used to classify individuals into discrete racial categories. This historical context raises two key questions about the citation trajectory of Lewontin 1972: first, what are the factors that caused an uptick in citations in the early 1990s after a prolonged period of bibliographic stagnation? Second, why did Lewontin 1972 ultimately emerge as the most iconic study of human genetic diversity from this era?

3. The influence of influential citations

We first considered the extent to which the citation trajectory of Lewontin 1972 was shaped by having accumulated citations in other widely cited scientific publications, which may have exponentially increased its exposure to other researchers. To

do this, we operationally define a publication as ‘influential’ if it has received over 1000 citations to date according to Semantic Scholar; by this metric, 11 highly influential papers and eight highly influential books/textbooks have cited Lewontin 1972. Seven of these were published in the 1972–1982 period, when there was an initial flurry of activity surrounding the topics of variance partitioning and classification in human population genetics; notably, these works referenced Lewontin’s empirical and statistical results but did not dwell on his sociological conclusions [20,34–39] (figure 1c). Most notably, the influential citing publications from this era include three widely read books that also set the stage for Lewontin’s prolonged feud with EO Wilson over sociobiology, as previously described by Segerstrale [40]: Lewontin’s [20] own *The genetic basis of evolutionary change*, an edited compilation of lecture notes, published in 1974; Wilson’s [35] *Sociobiology: a new synthesis*; and Gould’s [39] *The mismeasure of man* (note that the citation counts of these works sometimes drastically differ between Semantic Scholar and Google Scholar. For example, *The mismeasure of man* has been cited over 14 000 times according to Google Scholar but only approx. 4000 times according to Semantic Scholar).

No bibliographically influential citing publications appeared again until 1992, with Berry *et al.*’s [11] psychology textbook *Cross-cultural psychology: research and applications*.

Then in 1994, Bowcock *et al.* [41] replicated Lewontin's finding that most genetic variation exists within populations, using newly available polymorphic microsatellite data. The same year, Cavalli-Sforza *et al.* [42] published their landmark book, *The history and geography of human genes*. In 1996, Lande's [43] paper, 'Statistics and partitioning of species diversity, and similarity among multiple communities', again cited Lewontin 1972 for its contributions towards establishing a theoretical framework of variance partitioning within a species. The next two influential citing articles were published in 1997 and 1999 by the eminent sociologist and scholar of health disparities, Williams, and co-workers [12,44]. Seven other highly influential citing publications emerged during the 2000s, again turning the spotlight onto the technical aspects of Lewontin 1972 [13,45–49].

Most of these influential citations were unsurprising, in that they generally focussed on building upon and reassessing Lewontin's methods or reiterating his empirical results. The influential citations from Williams and co-workers [12,44], however, were an anomaly, not only because they came from the social sciences, but also because they shifted the focus onto Lewontin's interpretations, paraphrased as follows: *race is a gross indicator of distinctive social and individual histories and not a measure of biological distinctiveness*. While surveying the citing literature from the social sciences and the humanities, we came across another noteworthy paper, legal scholar Haney López's [50] 'The social construction of race: some observations on illusion, fabrication, and choice', published in the Harvard Civil Rights-Civil Liberties Law Review in 1994. This paper was reprinted a year later in the book *Critical race theory* [51], which is regarded as one of the foundational texts on the subject and has been cited over 3000 times, yet neither Google Scholar nor Semantic Scholar had indexed this book as having cited Lewontin 1972. In the light of these citations, we speculated that Lewontin's provocative conclusions may have been a unique feature of his paper that particularly appealed to other scholars in the social sciences and humanities, contributing to its iconic status.

To further quantify the importance of citations from the social sciences, we annotated each citing journal article with the inferred academic discipline based on the Scopus journal taxonomy from October 2020, which classifies articles into four top-level categories (biological sciences, physical sciences, health sciences and social sciences). We identified $N=1330$ citing articles for which these data could be retrieved and examined the changes in relative abundance of each discipline among this subset of citing articles over time. Although the majority of citing articles were published in life sciences journals over the paper's lifespan (figure 2*a,b*), we found that citing articles in social sciences journals did have the highest relative prevalence between 1985 and 1990 (figure 2*b*), lending some credence to our hypothesis that the social sciences provided a boost in bibliographic exposure to Lewontin 1972 at a time when citations in the life sciences literature were particularly sparse. Citing articles in the social sciences, however, never grew to be cumulatively more common than those in the life sciences, and it was clearly in the life sciences where Lewontin 1972 saw the biggest surge in citations in the last 20 years (figure 2*a*). Even so, it is worth noting that references in the social sciences tended to differ in *how* they cited Lewontin 1972. Using Semantic Scholar's 'citation intent' flag, which infers whether a given citation was part of the background, methodology or results,

we found that citing articles in the social sciences almost always (82% of papers where data are available) cited Lewontin 1972 as background, compared to citing articles in the life sciences, in which only 37.5% cited Lewontin 1972 as background and the majority (62.5%) of citations were found in the methods and/or results.

4. The influence of Lewontin's popularity as a science communicator

Given that Lewontin's [20] *The genetic basis of evolutionary change* was among the earliest influential citations, we also explored whether Lewontin had a tendency to self-cite his 1972 paper in his later works. We found only three other instances of self-citation, all published much later in his career: Lewontin & Hartl [52], where the authors discuss the usage of population genetics in forensic DNA typing (333 citations); Feldman & Lewontin [53], in a chapter titled 'Race, ancestry, and medicine' contributed to the book *Revisiting race in a genomic age* (54 citations); and Fujimura *et al.* [54], that argued the authors of a 2012 paper in the journal *Sociological Theory* had made claims 'based on fundamentally flawed interpretations of current genetic research' (67 citations). Although the relative paucity of self-citations for this paper indicate that it did not serve as the foundation of a major line of Lewontin's subsequent research (later in his career, he supposedly stated that he viewed the topic of genetics and race as 'completely uninteresting' [3]), the results and conclusions of Lewontin 1972 were undoubtedly foundational to his interpretation of human genetic diversity and its relevance to social and political issues. For example, in *The genetic basis of evolutionary change* (1974), Lewontin forcefully reiterates his position:

The taxonomic division of the human species into races places a completely disproportionate emphasis on a very small fraction of the total of human diversity. That scientists as well as nonscientists nevertheless continue to emphasize these genetically minor differences and find new 'scientific' justifications for doing so is an indication of the power of socioeconomically based ideology over the supposed objectivity of knowledge [20, p. 156].

It is also worth noting that during the 1982–1994 period when the bibliographic influence of Lewontin 1972 appeared to have stagnated, Lewontin published four of his most celebrated books, all of which touch on the topic of race and genetics and Lewontin's broader arguments against genetic essentialism, yet do not directly reference Lewontin 1972: *Human diversity* (1982) [55], *Not in our genes* (1984) [21], *The dialectical biologist* (1985) [17] and *Biology as ideology: the doctrine of DNA* (1992) [56]. Lewontin also defended the conclusions of his paper in other public-facing media outlets throughout his career, including televised interviews in 1975 [57] (an episode of PBS' NOVA programme that featured separate profiles of Lewontin and famous physicist Richard Feynman) and in 2003 [8]. Lewontin's books, radio/television appearances and his central role in the sensationalized debate over sociobiology in the 1970s–1980s may have served to keep Lewontin's earlier scholarly work salient in the minds of his peers and the next generation of academics.

5. Co-citation analysis

To search for additional clues about what may have sparked the influx of citations in the 1990s, we next examined which other papers were most commonly cited in the corpus of

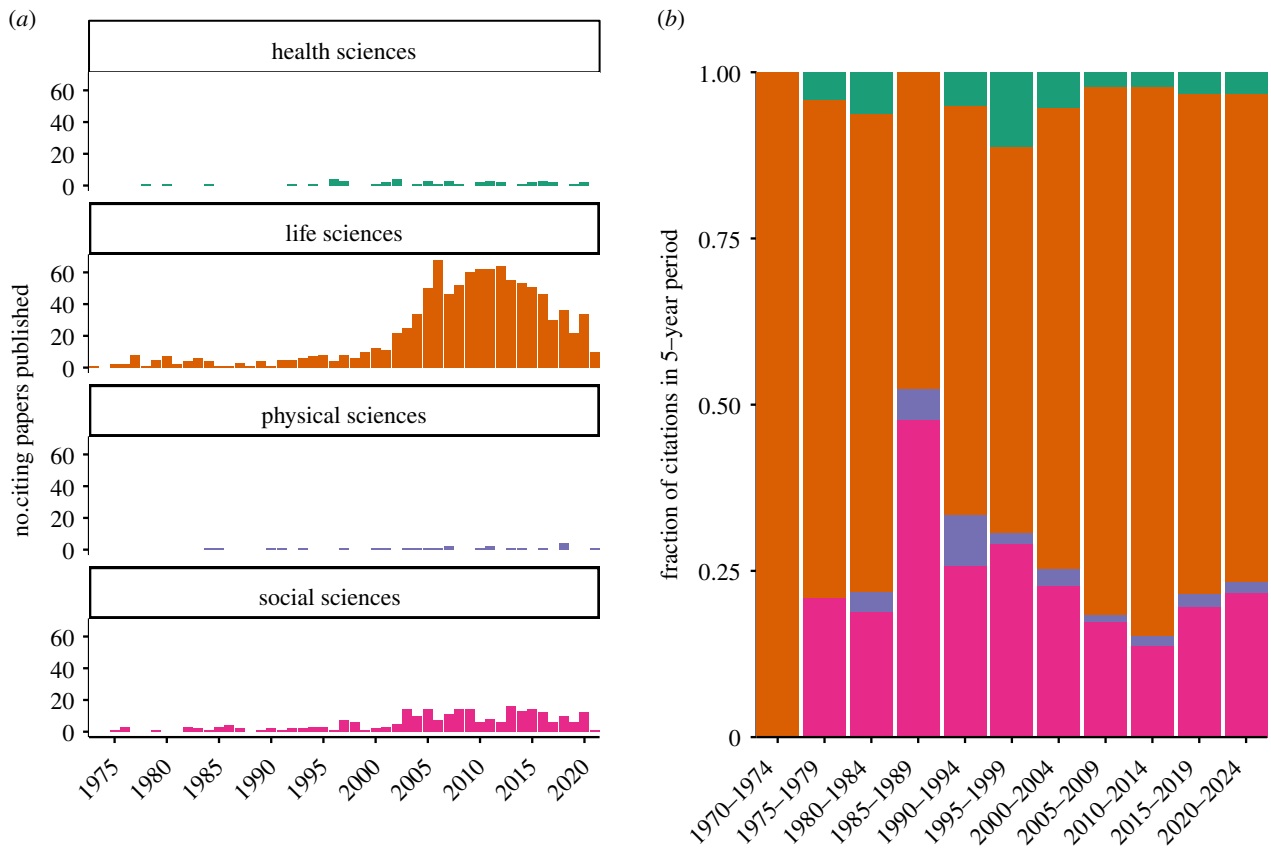


Figure 2. Breakdown of citation patterns for Lewontin 1972 according to the inferred research domain of the citing articles. (a) Total number of citing articles in a given year, stratified by research domain; (b) fraction of citing articles published in a given 5-year period within each of the four research domains.

research articles that referenced Lewontin 1972 (figure 3a). The most frequent co-citation by far—including every year since 2009—was Nei's [34] 'Analysis of gene diversity in subdivided populations', which was also the most influential citing article (figure 1c). In addition, three of the 15 most common co-citations are also celebrated population genetics papers from the 1960s and 1970s: Kimura & Crow's [58] 'The number of alleles that can be maintained in a finite population', Nei's [59] 'Genetic distance between populations' and Nei's [60] 'Estimation of average heterozygosity and genetic distance from a small number of individuals'. When we compared the citation trajectories of these four papers, we found that they closely mirrored that of Lewontin 1972 (figure 3b), which implies that the underlying factors that contributed most strongly to the secondary surge in citations of Lewontin 1972 were likely common to other important population genetics papers from that era.

As described in the previous section, one of the first influential citing papers in the 1990s was Bowcock *et al.* [41], which, in 1994, replicated Lewontin's variance partitioning results using a novel type of genetic marker, polymorphic microsatellites (another publication in 1994 would do the same using restriction fragment length polymorphisms (RFLPs) [61], though this paper has been less widely cited). Our co-citation analysis revealed that Lewontin 1972 was frequently cited alongside other articles describing novel genetic markers that emerged in the 1990s: Williams *et al.*'s [62] 'DNA polymorphisms amplified by arbitrary primers are useful as genetic markers' which introduced a genotyping technique based on Random Amplified Polymorphic DNA (RAPD) markers, and Vos *et al.*'s [63] 'AFLP [Amplified Fragment Length Polymorphisms]: a new technique for DNA fingerprinting' (figure 3a). When first introduced in 1990,

RAPD markers offered several advantages over the prevailing source of polymorphic data, RFLPs, and were later described as being particularly useful in advancing empirical population genetics research [64]. This suggests that the renewed interest in Lewontin 1972 (and other papers from that era) was driven by the rapid evolution of new genotyping technologies in the late 1980s and early 1990s. This is all but confirmed by several of the remaining articles we found to be most frequently co-cited with Lewontin 1972, which focused on the same question of quantifying genetic variation within and between populations: Excoffier *et al.* [65] used mitochondrial RFLP data, Barbujani *et al.* [66] used autosomal RFLP data and Rosenberg *et al.* [46] used autosomal microsatellite data (figure 3a). Other notable co-citations include several popular population genetic software programs first introduced in the 1990s and 2000s (genalex [67], NTSYS-pc [68], STRUCTURE [69] and PopGene [70]), which suggests that rapid advances in computational power may have further precipitated and enabled interest in revisiting the early works of Lewontin, Nei, Kimura and others using new datasets and computational methods. The other two top co-citations we identified were Mantel's [71] 'The detection of disease clustering and a generalized regression approach', which introduced a statistical test that came to be widely adopted in population genetics for evaluating population structure [72], and Hamrick *et al.* [73] 'Allozyme diversity in plant species', underscoring that the relevance of Lewontin's work extended far beyond the human species.

6. What happened to Nei and Roychoudhury?

One outstanding question that these citation and co-citation patterns still do not explain is why Lewontin's paper went

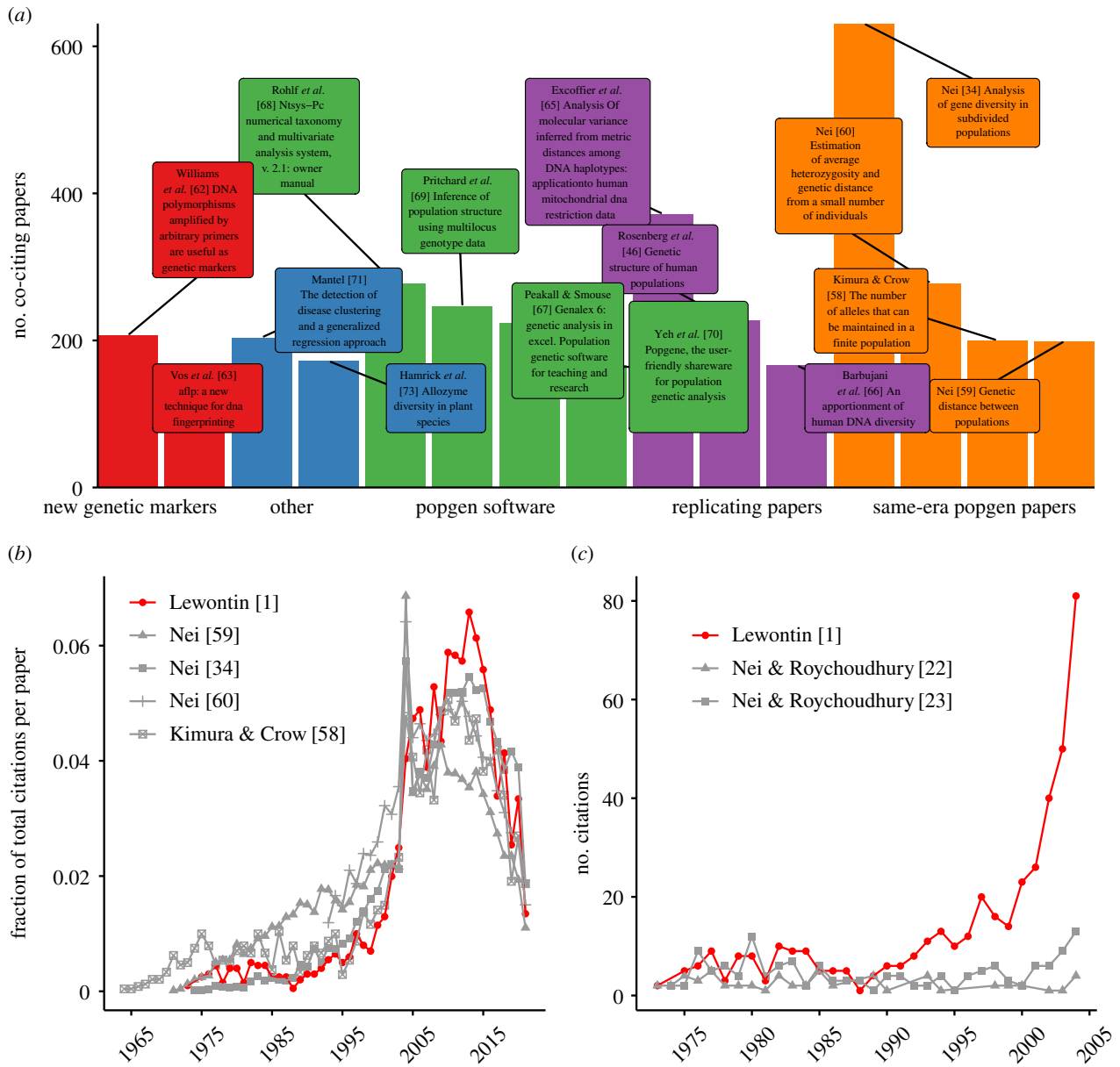


Figure 3. (a) Co-citation frequencies over time for the 15 papers most commonly referenced alongside Lewontin 1972. Data prior to 1995 are not shown due to much lower overall citations of Lewontin 1972. (b) Citation trajectories of landmark population genetics papers (each published before 1980) that were among the most common co-citations of Lewontin 1972. (c) The citation trajectory of Lewontin 1972 compared to topically similar papers published by Nei and Roychoudhury [22,23] that reported similar results.

on to amass thousands of citations in the last 30 years, yet the comparable studies published by Nei & Roychoudhury [22,32,74] did not. Though the papers of Nei and Roychoudhury never received enough citations to be counted among the most frequently co-cited with Lewontin 1972, we confirmed that they are co-cited in several of the most highly cited publications we identified, including Nei [34], Bowcock *et al.* [41], Cavalli-Sforza *et al.* [42] and Haney López [50]. In addition, the citation trajectories of Nei's and Roychoudhury's papers from 1972 to 1974 are very similar to that of Lewontin until the early 1990s (figure 3c). Furthermore, as early as the mid-1980s (and perhaps even earlier), both Lewontin 1972 and Nei and Roychoudhury's work had been invoked in arguments rejecting the biological basis of race [40,75]. These bibliographic breadcrumbs suggest that for the first 20 years, Lewontin 1972 did not have broader exposure to fields outside of population genetics, nor was it viewed as having bibliographic precedence in the literature, nor did it occupy an exclusive role in shaping the scientific consensus

that race is not taxonomically meaningful. However, the citation trajectories shown in figure 3c suggest that the early 1990s were an inflection point at which Lewontin went on to be cited much more frequently than Nei and Roychoudhury, and indeed, only two of the nine influential papers published after 1994 cited both Lewontin 1972 and Nei and Roychoudhury.

To resolve this question, we turned to several key papers in the corpus of citing literature that contrasted the rhetoric of Lewontin 1972 with that of Nei's and Roychoudhury's publications. As described by Haney López [50], the two diverged dramatically in their interpretations of the variance partitioning results:

Lewontin argued that biologists should abandon all talk of biological races [...]. Nei and Roychoudhury agree that talk of biological races should be abandoned, but point out that there remain statistically significant differences between smaller population groups that justify the continued scientific division of humans by gene type [50, p. 12].

Brown and Armelagos would later argue that the conclusions of Nei and Roychoudhury were logically inconsistent with their results:

Interestingly, despite these very low figures [of between-population variance], [Nei and Roychoudhury] went on to discuss ‘the pattern of evolution of the three major races’ (p. 11). This speaks to the logical disconnect shown by many researchers who simultaneously prove the irrelevance of genetic race and then proceed to discuss the genetic evolution of races [4, p. 36].

In Stephanie Malia Fullerton’s retrospective analysis [76] of Appiah’s [75] celebrated essay on the philosophy of race, ‘The uncompleted argument: Du Bois and the Illusion of Race’, she explains how Appiah names Nei & Roychoudhury [32] as paramount evidence for rejecting the biological basis for the concept of human races (in a footnote, Fullerton acknowledges Lewontin 1972 as being one of a handful of scientific papers that was ‘publicly important much earlier’, but does not elaborate on what differentiated Lewontin’s conclusions from those of Nei and Roychoudhury). Even so, Appiah was aware of the fact that Nei and Roychoudhury still essentially held the view that race was taxonomically meaningful:

[Appiah] also acknowledged that the geneticists he cited [Nei and Roychoudhury] were those who ‘believe in human races’ but disputed their claim that their data ‘shows the existence of a biological basis for the classification of human races’ [76, p. 38].

Haney López’s [50, p. 13] paper even went so far as to conclude that Nei and Roychoudhury ‘reflexively fall into the comfortable habit of White supremacy in science’, a statement motivated by two central tenets of Critical Race Theory: first, that the concept of ‘biological race’ is rooted in systems of oppression and racism that serve to advance the interests of white people at the expense of people of colour; second, that ‘[races] are not objective, inherent or fixed, they correspond to no biological or genetic reality; races are categories that society invents, manipulates or retires when convenient’ [51, p. 7]. Although humanist scholars such as Appiah and Haney López did not take issue with Nei and Roychoudhury’s technical variance partitioning results, they were not convinced that modest between-population variation was sufficient to support the weight accorded to racial categories within society. Ultimately, Lewontin’s blunt interpretations of his variance partitioning results appear to have proved much more attractive to social scientists weighing the evidence for any biological basis of race, and these scholars’ citations likely helped boost Lewontin 1972 to twenty-first-century prominence.

7. The impact of the human genome project

The Human Genome Project (HGP) was one of the most massive scientific endeavours of the twentieth century, and it catalysed an extraordinary increase in the amount of data available for research on human genetic variation. Given the foundational importance of Lewontin’s work towards the interpretation of genomic data, we considered the possibility that the HGP’s architects were familiar with Lewontin 1972 and helped propel it into the spotlight. Chronologically, this seems like a plausible explanation: the HGP was announced in 1990 and the completion of the first drafts of the human genome were published in 2001; these 2 years roughly align with marked increases in citations of Lewontin

1972 (figure 1b). However, although planning documents for the HGP mentioned the importance of the project’s ethical, legal and social impact (including how genomic data might be misused to ‘advance eugenics or prejudicial stereotypes’ [77]), we found no evidence in the bibliographic record that Lewontin 1972 was ever explicitly referenced by the organizations and key personnel involved in initiating the HGP, nor was it referenced in any of the papers or commentaries appearing in special issues of *Nature* and *Science* announcing the completed draft sequences of the human genome in 2001.

Although we found no evidence that the HGP had a direct bibliographic connection to Lewontin 1972, it likely played a role in popularizing the sound bite ‘there is more genetic variation within populations than between populations’, which is deeply intertwined with Lewontin’s legacy today [9]. Earlier literature suggests Lewontin himself was circulating versions of this sound bite in popular media as early as the mid-1970s [40], but it seems that, over time, the aphorism came to be disconnected from Lewontin as its originator. This disconnect is plainly apparent in Edwards’ [78] critique of Lewontin 1972. Edwards begins his paper by mentioning that when the first draft of the human genome was published in *Nature* in 2001, the print version of the journal shipped with a compact disc of educational materials pertaining to human genomics research (the materials from this CD, titled ‘Understanding the Human Genome Project’, are currently hosted online by the National Human Genome Research Institute [7]). These materials include a module about human genetic variation that stated ‘there is more genetic variation within populations than between populations’, but without citation. Edwards took issue with how this statement had been used to ‘play down the genetical differences among human populations ... usually without reference’ [78], but despite his personal and scholarly familiarity with Lewontin, he was entirely unaware of Lewontin’s past association with this statement [79]. Edwards’ eventual discovery that the statement could be traced back to Lewontin 1972 led him to effectively attribute the paper as an uncited source for both the HGP and a backlog of highly influential news media that had indirectly popularized the paper’s conclusions. One such example we identified was a New York Times article from 2000 in which the author briefly summarizes the within- and between-population variance partitioning results (which at this point had been replicated by multiple research groups) and contains quotes from several prominent scientists (including those involved with the HGP) indicating their familiarity with these results, but Lewontin is never mentioned [80].

Though Lewontin 1972 was being cited more than ever by the turn of the twenty-first century (figure 1b), these anecdotes suggest that the results of the study had become so widely known and accepted that attribution was not always considered necessary. Paradoxically, the HGP’s singular focus on a single genome had, from the onset, incited researchers to wonder how much was omitted from that narrow view of human genetics [81]. By giving scientists more of a reason to think about human genetic variation, the HGP may have prompted them to rediscover and bibliographically reassert Lewontin 1972 as they sought to augment our understanding of human diversity with data from more individuals and populations. Ironically, Lewontin was highly critical of the ideological motivations and rationale of the HGP and published several essays in the *New York Review of Books*, which were ultimately compiled in his 2000

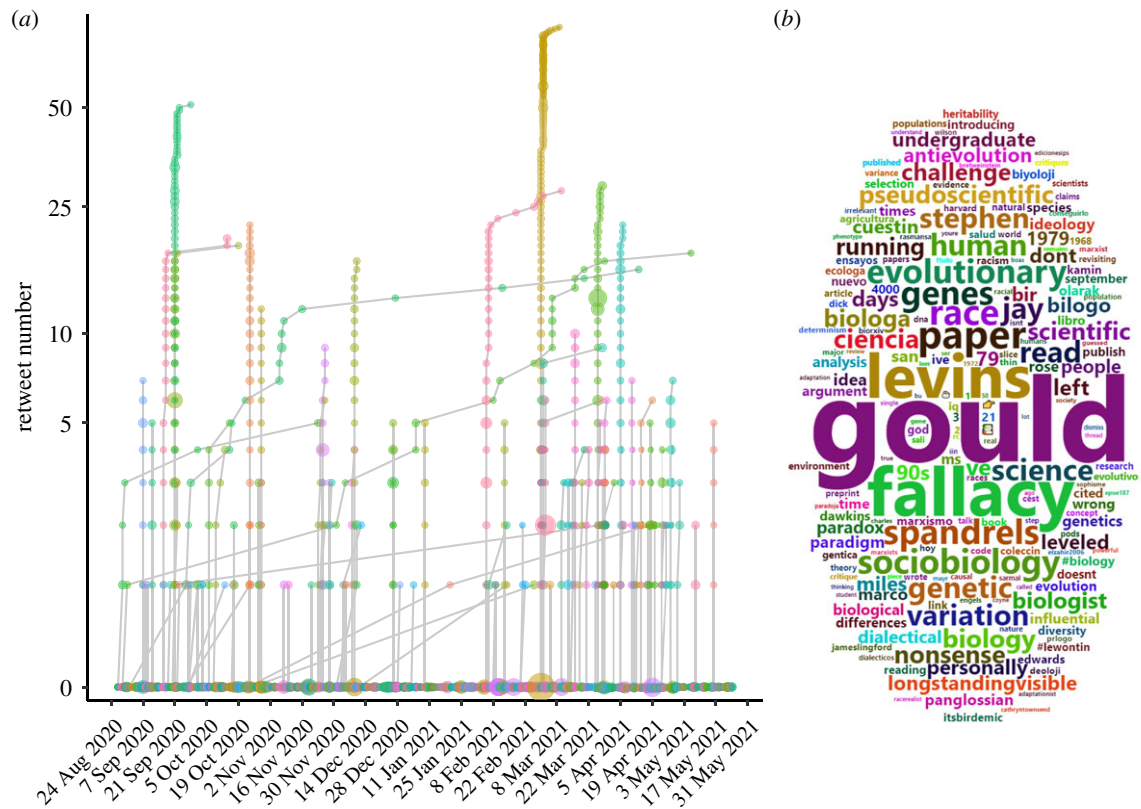


Figure 4. Summarizing nine months of Twitter data for tweets referencing ‘Lewontin’. (a) Timeline of tweets and retweets in our dataset. Each original tweet is shown as a point along the x-axis. For tweets that were retweeted, the retweet trajectories are indicated by linked line segments growing along the y-axis as the number of retweets increases over time. The size of each point indicates the number of followers of the corresponding user. (b) Wordcloud representing the 200 most frequent words found among the 2507 tweets collected (including retweets).

book, *It ain't necessarily so: the dream of the human genome and other illusions* [82].

8. Standard altmetric indicators of Lewontin 1972

Citations of Lewontin 1972 in the academic literature have declined precipitously in the last 5 years (figure 1b), rendering traditional bibliometrics relatively ineffectual for understanding the ongoing impacts of the paper. However, the recent rise of social media as a vehicle for discussing scholarly research has enabled new paradigms for documenting and measuring how papers are shared and discussed [83]. The Altmetric Attention Score, which aims to quantify the attention received by scholarly works on social media, news media, blogs, Wikipedia and other non-traditional sources of citations, is one of the most popular altmetric indicators used by the research community. As of 15 June 2021, Lewontin’s 1972 paper had an Altmetric Attention Score of 85, with only 17 tweets from 17 unique users directly referencing the paper [84]. By contrast, Edwards’ [78] critique had received nearly 10-fold more directly referencing tweets (145 tweets from 102 unique users, as of 15 June 2021) and had a higher Altmetric Attention Score of 109 [85] despite having only received approximately one-tenth as many literature citations as Lewontin 1972 (318 citations of [78], compared to 3076 citations of [1], per Google Scholar). Although its relatively modest Altmetric Attention Score and lack of ‘indexable’ Twitter citations (i.e. tweets including a DOI or URL link to an online version of the paper) might create

the illusion that Lewontin 1972 has little impact within the rapidly changing landscape of social media, a closer examination shows that it is widely and actively discussed on Twitter on a daily basis, underscoring a prolonged cultural influence that cannot be directly ascertained from traditional bibliometrics or even standard altmetric indicators.

9. The impact of Lewontin 1972 as revealed by ‘dark citations’ on Twitter

Using the Twitter API, we collected all tweets (and any subsequent retweets) containing the word ‘Lewontin’ that were posted over a nine-month period from 27 August 2020 to 25 May 2021, resulting in a collection of 2659 data points (tweets/retweets), an average of 9.8 tweets/retweets per day during this period. We hypothesized that these tweets would contain a number of ‘dark citations’ to Lewontin 1972, defined by Jensen [18] to be references to scholarly works that do not necessarily include traceable links such as DOIs or URLs. After excluding 152 tweets/retweets referencing the R.C. Lewontin Early Award from the Society for the Study of Evolution, which was soliciting nominations during this period, we confirmed this dataset did not contain any additional off-topic tweets, e.g. from (or in response to) unrelated users with the surname ‘Lewontin’. A timeline of these tweets and their retweets during this period is shown in figure 4a. This timeline paints a picture of a steady, ongoing conversation about Lewontin’s work rather than a flurry of activity surrounding any specific controversies or events. Most tweets (1636/2507; 65%) were original unique

tweets, not retweets; moreover, the 2507 tweets/retweets came from 1589 unique users, demonstrating that these tweets are not simply the output of a few individuals with a particularly strong interest in Lewontin. Of the 1636 original tweets, 1381 were quote-tweets or replies to other tweets, suggesting that the majority of Twitter references to Lewontin primarily arise in debates and conversational contexts. Though most (1057/1636) of the original tweets were in English, the following languages each accounted for at least 10 tweets in our dataset: German, Spanish, French, Italian, Portuguese and Turkish. Furthermore, we observed tweets in Arabic, Danish, Japanese, Norwegian, Russian and Swedish, indicating a linguistically and likely geographically diverse audience. The most frequently used words present in this dataset of tweets are summarized as a wordcloud in figure 4b, which shows many of the most common words in this dataset refer to Lewontin's other highly cited works with coauthors: 512 include 'Gould', 269 include 'Levins' and 91 include 'Kamin' or 'Rose', referring to coauthors Stephen Jay Gould [19], Richard Levins [17], Steven Rose and Leon Kamin [21].

The most frequently used word in this dataset that is not a coauthor's surname is 'fallacy', found in 10% (243) of the 2507 tweets/retweets. This refers to the phrase 'Lewontin's Fallacy', popularized by Edwards' critique of Lewontin 1972 [78]. A broader search for keywords relevant to Lewontin 1972 within these tweets revealed that 21.4% of the tweets (536/2507, or approx. 2 tweets per day during the nine-month data collection period) include the words 'fallacy', 'diversity', 'race', 'racial', 'racism', 'racist', '15%', '85%' or 'variation' (we included '15%' and '85%' in these filtering criteria because they refer specifically to the percentages of human genetic diversity Lewontin attributed to between-group and within-population differences, respectively [1]). By contrast, citations of Lewontin 1972 account for only 4% (3076/75 637) of Lewontin's total citations, according to Google Scholar. This suggests that Lewontin 1972 has an extraordinarily outsized influence in the social media ecosystem, relative to Lewontin's broader body of work.

We also applied our recently developed social media audience segmentation method [86] to categorize the users that engaged in conversations about Lewontin 1972 on Twitter over the nine-month data collection period. Briefly, we identified the followers of each of the 303 unique users whose tweets/retweets contained keywords indicating a specific reference to Lewontin 1972 (described above) then applied a statistical model to identify the most common co-occurring words in the bios of each focal user's followers as an indicator of the network(s) each user is affiliated with. As described in Carlson & Harris [86], we interpret the properties inferred from each focal user's followers as characteristic of that focal user, according to the principle of network homophily. The results of this audience segmentation are presented in figure 5. Based on the presence of particular keywords in the inferred audience topics (using the criteria described in [86]), we estimate that 44.2% of users in this dataset are primarily affiliated with academic communities, including medicine (sector 4), philosophy/psychology (sector 7), ecology (sector 8), genomics/bioinformatics (sectors 10 and 11) and evolutionary biology (sector 12), roughly mirroring the distribution of literature citations across life sciences, health sciences and social sciences as portrayed in figure 2 (though none of the

academic audience sectors appear to align with fields in the physical sciences).

Two audience sectors appeared to be primarily characterized by the industry/occupation of their constituent users: software engineering/development (sector 1) and business and marketing (sector 6), which together accounted for 7.4% of users in the dataset. The remaining audience sectors, accounting for 48.4% of users in the dataset, were primarily characterized by political ideology/affiliation. Most (52%) of these politically affiliated users aligned with right-wing American politics (sector 3) and likely fell along a spectrum ranging from mainstream Republican supporters (evidenced by the keywords 'conservative', 'trump' and the unicode symbol 'US', which is displayed as the American flag emoji on Twitter) all the way to adherents of far-right identitarian ideologies (evidenced by the keywords 'white' and 'nationalist'). Sector 9 was mainly represented by French keywords and political terminology; the presence of the hashtag '#team-patriotes' in the top keywords (which appears to be popular among French Twitter users who support far-right political candidates) and at least one tweet originating from a prominent French white nationalist (see Discussion) suggests that users associated with this sector also tend to align with far-right ideologies. Conversely, sector 2 appeared to capture users aligned along the spectrum of left-wing politics, evidenced by the keywords 'socialist', 'feminist' and '#blacklivesmatter', accounting for approximately 20% of the politically affiliated audience. The remaining audience sector, sector 5, appeared to be a 'generalist' category of users interested in politics among a variety of other topics (including science, culture, technology and history), but the top keywords do not indicate a specific political ideology and closer qualitative examination of the users affiliated with this sector suggests they identify across the political spectrum.

10. Discussion

Our analysis of the bibliometric and altmetric indicators of Lewontin 1972 provides an important backdrop for understanding and contextualizing the significance of its scientific and cultural impacts. The citation trajectory shows that references to Lewontin 1972 were sustained at a modest rate through the 1970s and early 1980s (peaking at around 10 citations per year in 1982), then tapered off into the late 1980s, only to be followed by a surge of bibliographic attention beginning in the early 1990s and extending well into the post-genomic era, peaking at over 130 citations per year in 2013, over 40 years after its initial publication. The events and publications surrounding Lewontin 1972 that we analysed here are summarized on a timeline in figure 6.

As revealed by our analysis of influential citing papers and co-citations, there were several trends and events that coalesced in the early 1990s (and in 1994, specifically) to propel Lewontin 1972 to its current iconic status. First, the field of empirical population genetics research entered into somewhat of a renaissance in the early 1990s [87]. By the early 1980s, the initial flurry of debate over Lewontin's results had mostly subsided, and data from classical genetic markers (e.g. blood groups, proteins) had been extensively mined. A decade later, the emergence of novel genetic markers that could be easily assayed genome-wide (e.g. RFLPs, microsatellites), along with increasingly sophisticated computational

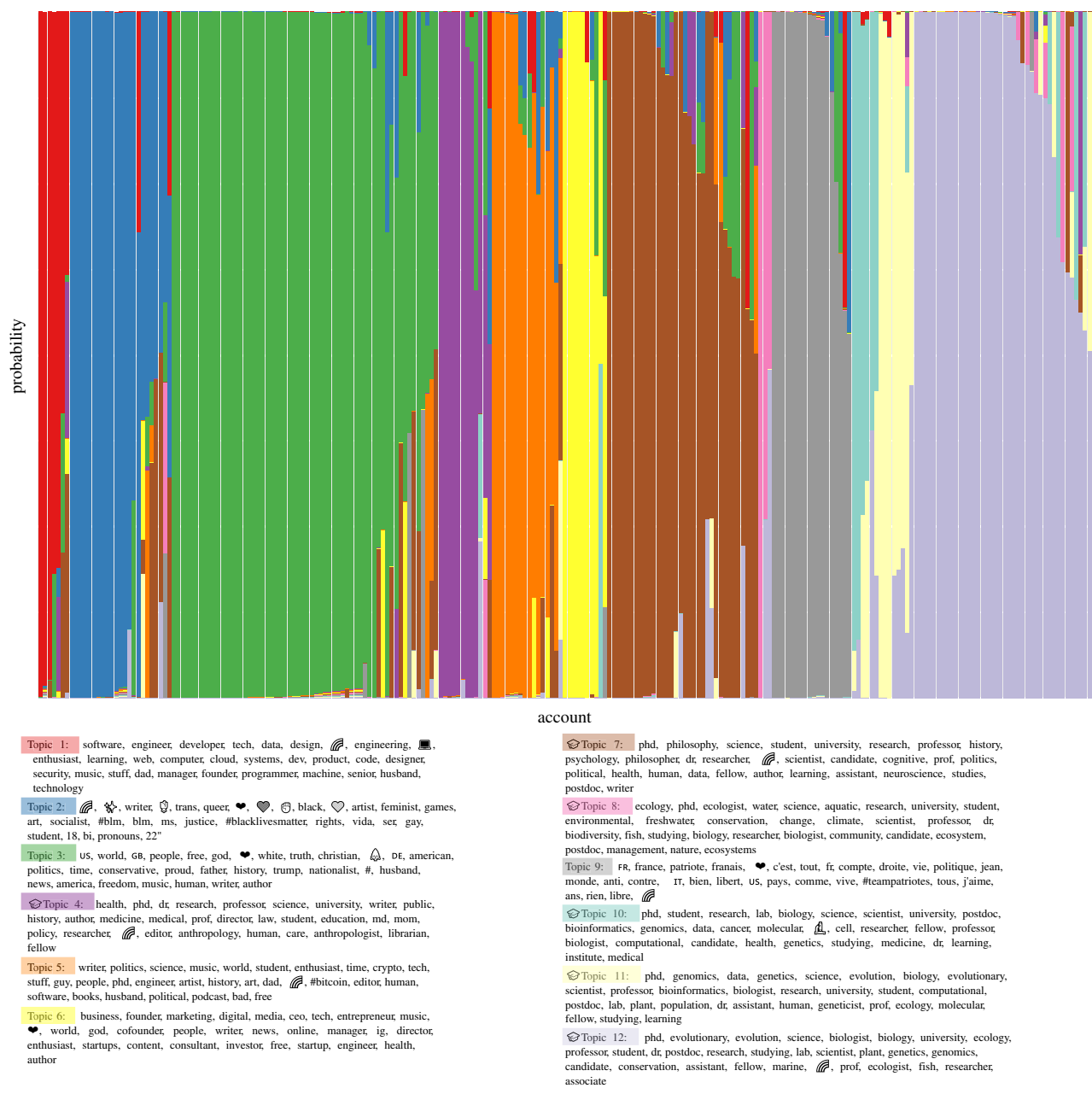


Figure 5. Twitter audience segmentation analysis for users indirectly referencing Lewontin 1972. Each unique user that tweeted about that paper is represented by a vertical stack of 12 coloured bars that represent the user’s estimated membership probability in each of the 12 inferred audience sectors. The top 30 keywords, hashtags or emoji associated with each audience sector are shown in the legend. Audience sectors inferred to correspond to academic communities are indicated with a 🧑‍🎓 emoji at the beginning of the list of associated keywords.

tools for analysing these data, made it possible to empirically revisit decades-old theories, hypotheses and results in the field. With respect to Lewontin 1972, the fruits of this renaissance were first evident in two 1994 studies that replicated Lewontin’s results: Dean *et al.* [61], which analysed RFLP data, and Bowcock *et al.* [41], which analysed microsatellite data. By this point, Lewontin 1972 was both well-established in the population genetics literature and increasingly recognized for its contributions to the scientific consensus rejecting the concept of biological races, in large part due to Lewontin’s books and media appearances throughout the 1970s and 1980s [40]. In the same year these replication studies were published, two other highly influential works would cite Lewontin 1972: Cavalli-Sforza *et al.* published their celebrated book, *The history and geography of human genes* (later described as ‘the great synthesis of genetic data

with historical, archaeological and linguistic information’ [88]), and Haney López published a paper invoking Lewontin’s results as foundational to Critical Race Theory, which was republished in one of the most important books on the topic [51] the following year.

In 1994, Herrnstein & Murray [89] also published their book, *The bell curve*, which immediately attracted widespread controversy over its claims that racial differences in IQ were due to innate genetic differences between races. Though Lewontin had sparred with Herrnstein decades earlier about the heritability of IQ [90] (and by this time had established himself as a popular and outspoken opponent to the genetic essentialism espoused by Herrnstein & Murray [21]), Lewontin 1972 was not cited in *The bell curve*, and the earliest publication in which Lewontin mentioned *The bell curve* (that we are aware of), was not until his 2001 book, *It*

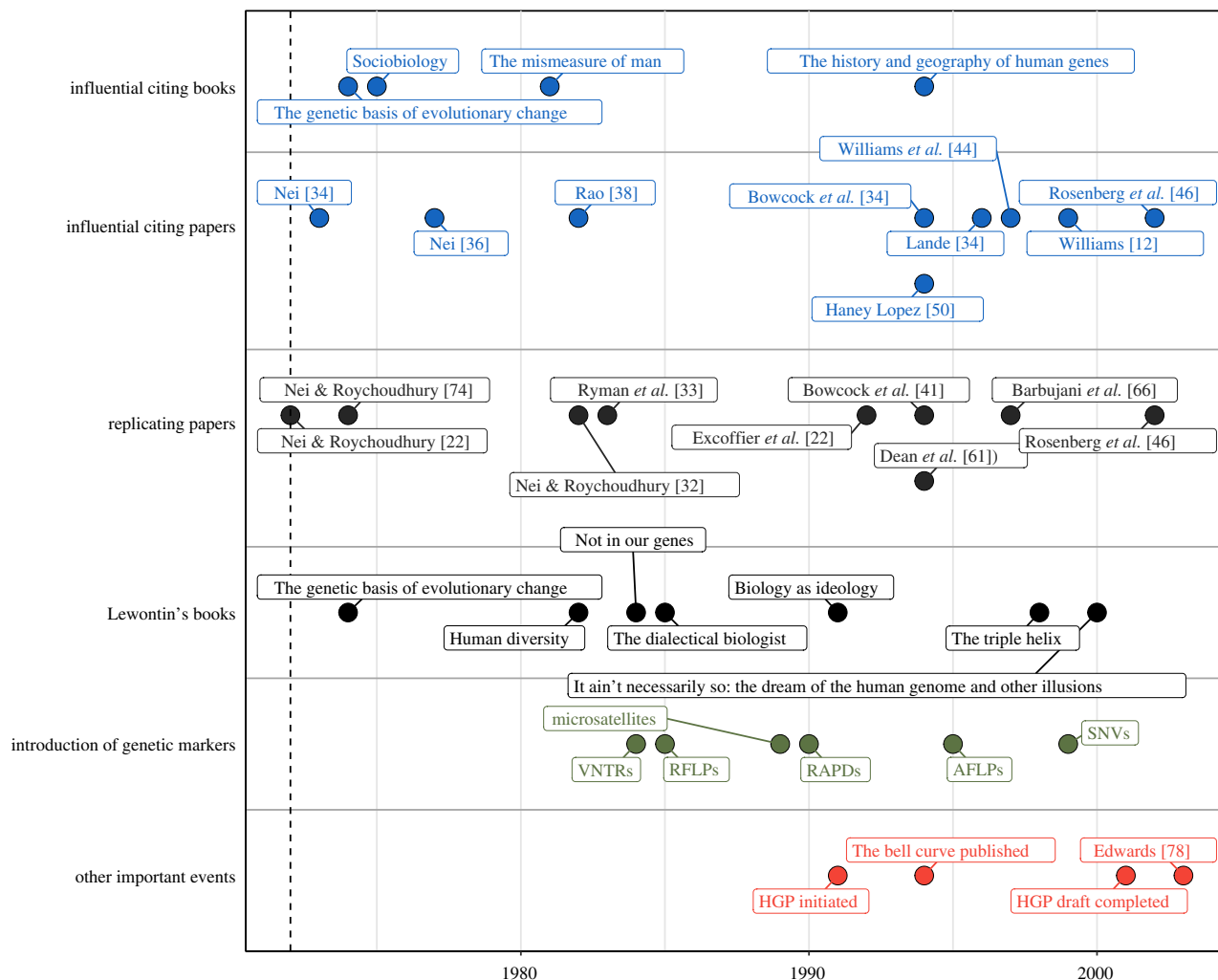


Figure 6. Timeline summarizing the relevant events, publications, citations and co-citations surrounding and involving the bibliographic history of Lewontin 1972, up to 2005.

ain't necessarily so: the dream of the human genome and other illusions [82], in which he describes *The bell curve* as biological determinism's 'most famous literary incident'.

Even though there was no bibliographic connection between *The bell curve* and Lewontin 1972, nor did Lewontin appear to lead the charge in attacking its narrative of biological determinism (at least in his published works), the results and conclusions of Lewontin 1972 were repeatedly invoked by Herrnstein and Murray's legion of critics. In 1995, philosopher Ned Block wrote a scathing critique of *The bell curve* that mentions the results of Lewontin 1972 as being 'widely accepted by all sides' (note that Block cites Lewontin's book *Human diversity* as the source of these results rather than the 1972 paper) and suggested Murray had 'pathetically misunderstood' the technical nuances of variance partitioning [91]. In 1996, Gould's *The mismeasure of man* (which cited Lewontin 1972 when it was originally published in 1981) was revised and republished for its fifteenth anniversary with a new cover prominently proclaiming it was 'the definitive refutation to the argument of *The bell curve*'. In the introduction to this revised edition, Gould explicitly positions himself and Lewontin as the figureheads of opposition to Herrnstein, Murray and Richard Jensen [39]. The specific citation of Lewontin 1972 as it appears in Gould's *The mismeasure of man* was even mentioned in a 1994 letter to the editor in the *Washington Post* by a reader critical of *The bell curve*, demonstrating that, by this time, lay audiences were already well-

versed in applying the conclusions of Lewontin 1972 to refute scientific racism and genetic essentialism [92] (perhaps an outgrowth of Lewontin's central and highly publicized role in criticizing sociobiology throughout the preceding two decades [40]). In a 2005 retrospective op-ed about the impact of *The bell curve*, Charles Murray himself would invoke Lewontin 1972 as being directly relevant to this controversy, stating incorrectly that 'Richard Lewontin originated the idea of race as a social construct in 1972' [93] (in fact, scholarship on the social construction of race dates at least back to 1897 with Du Bois's [94] 'The conservation of races' and further popularized by anthropologist Franz Boas throughout the early twentieth century [95]).

Though we are unable to put any numbers to its effect, *The bell curve* undoubtedly played a role in popularizing Lewontin 1972 in the 1990s. We speculate that the prominence of *The bell curve* controversy may have been a motivating factor for the scholars who were amplifying Lewontin's results and ideas around this time, across both the life sciences and social sciences. Anyone troubled by the impact of the views expressed in *The bell curve*, as well as the associated representation of the field of genetics, may have reacted by scrutinizing the logical consistency of other publications they saw as having the potential to fan its flames (particularly Nei and Roychoudhury's works [40,50,75]), thereby contributing to the bibliometric stagnation of Nei and Roychoudhury's papers and the accelerating citation of Lewontin 1972. Similarly, although we found no

explicit citations of Lewontin 1972 in HGP-related publications and communications, the architects of this project may have been motivated to amplify the ‘sound bite’ of the paper because of its implicit endorsement of a single human reference genome as representative of all humanity.

Though literature citations of Lewontin 1972 have tapered off in the last 5 years, the Twitter data we analysed showcases how the study has persisted in the academic and public discourse, accumulating roughly two tweets per day, with approximately equal attention from academic and non-academic audiences. Our analysis differs from most other altmetric studies because we primarily focus on ‘dark citations’ found on social media (i.e. references to the study that do not directly link to a DOI or URL and thus are not tracked by Altmetric and other altmetrics data brokers). The concept of ‘dark citations’ has been previously used to describe the practice of referencing prior published work in a journal article without an explicit citation [18], but it is perhaps even more relevant for altmetrics where much of the social media attention surrounding a paper takes place in colloquial threaded discussions. Such data are arguably a more salient indicator of a paper’s societal/cultural impact than standard bibliometric/altmetric indicators because they demonstrate the degree to which a paper’s results and conclusions have become embedded in the public mindset, to the extent that directly referencing the paper is superfluous. In the case of Lewontin 1972, our analysis shows how the paper continues to be embroiled in the cultural reckoning of defining and applying the concept of human races.

We acknowledge that our search criteria for these ‘dark citations’ were not exhaustive; tweets containing some variation of the phrase ‘there is more genetic variation within populations than between populations’ were excluded unless they explicitly included the word ‘Lewontin’. This may have biased our dataset to more thoroughly sample tweets that were critical of the study and/or Lewontin himself: because the phrase ‘Lewontin’s Fallacy’ is popular among Lewontin’s critics and detractors, their tweets would have been included, whereas those who supportively repeat the ‘sound bite’ of the paper may be entirely unaware of who Lewontin is and his role in popularizing this interpretation. Ironically, our data shows that, among non-academic audiences on social media, the groups most ideologically opposed to Lewontin’s claims play a significant role in maintaining the connection between Lewontin and the broader concept that ‘there is more genetic variation within a population than between populations’.

It should be noted that the tweets in our dataset that mention ‘Lewontin’s Fallacy’ can be construed as indirect references to Edwards’ critique [78] rather than Lewontin’s original paper [1]. Though this distinction has no practical bearing on our analyses, it may be a useful one to make in understanding the motivations of those who continue to keep this paper in the spotlight. Many critical tweets in our dataset appear to invoke ‘Lewontin’s fallacy’ simply as a rhetorical cudgel in an attempt to dismiss an opposing argument as logically invalid. For example, one Twitter user states: ‘This fallacy is so well known in science that it has a name: Lewontin’s fallacy’. Similar tautological claims are echoed by ideological critics in the scientific literature; for example, Warne *et al.* [96] states ‘[a fallacious claim that genetic similarity among humans negates phenotypic differences] is so common in the biological and social science

literature that it even has its own name: *Lewontin’s fallacy*, named for a biologist who popularized it’. Such statements lionize Edwards’ critique as the authoritative interpretation of Lewontin’s results and paint Lewontin as a solitary proponent of this claim, despite more contemporary research that has largely vindicated Lewontin’s interpretation and demonstrated that Lewontin did not, in fact, commit ‘Lewontin’s fallacy’ (e.g. [6,66,97–102]). Moreover, we find that the Twitter users who vehemently oppose the conclusions of Lewontin 1972 often have significant overlap with extreme far-right political communities, underscoring how rejection of Lewontin’s interpretation has become a tenet of white nationalist ideology.

One such individual who appeared in our dataset is Renaud Camus, a prominent white nationalist French writer who gained notoriety for coining the term ‘Great replacement’, a conspiracy theory that postulates white European populations are being demographically and culturally replaced by non-white immigrants through policies enacted by ‘the global elites’ [103]. Camus’ tweet states *Jamais la Science ne se sera montrée plus servée*, translating to ‘Science will never have shown itself to be more useful’, and quotes another user who wrote:

Indeed, it was in 1972 that the geneticist Richard Lewontin published the article ‘The Apportionment of Human Diversity’, which should be a landmark in the construction of anti-racist dogma. (Translated from French by Google Translate software.)

This conversation traces back to an earlier tweet posted by Camus, where he claims:

In 1976 the Haby law [legislation introducing educational reforms in France]—by inaugurating the single college—abolished classes, eradicated the cultivated class and therefore culture; at the same time the dogma of the inexistence of races eradicated the white race and therefore Western civilization. (Translated from French by Vivian Link.)

Taken in the context of this thread, we interpret Camus’ original tweet to insinuate that he views Lewontin as the architect of an ‘anti-racist dogma’ that put into motion the ‘eradication’ of the white race. Given that Camus’ Great Replacement theory is widely regarded as a core belief among white nationalists [104] and has been explicitly cited in manifestos written by the perpetrators of mass shootings in Christchurch, New Zealand, Pittsburgh, Pennsylvania and El Paso, Texas [105], his familiarity with Lewontin’s paper is an alarming reminder that scientific research does not occur in a political vacuum. Just as Lewontin was acutely aware of the political minefield surrounding human genetics research and spoke out against its misuse and misappropriation, our findings make the case that scholars in the fields that have been shaped by Lewontin 1972 bear a moral and ethical responsibility to do the same.

Data accessibility. Code and data underlying these analyses are available at <https://github.com/carjed/lewonlin>.

Authors’ contributions. J.C.: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing—original draft, writing—review and editing; K.H.: conceptualization, funding acquisition, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. K.H. acknowledges support from NIH grant no. 1R35GM133428-01, a Searle Scholarship, a Sloan Research Fellowship, a Pew Biomedical Scholarship and a Burroughs Wellcome Fund Career Award at the Scientific Interface.

References

- Lewontin RC. 1972 The apportionment of human diversity. In *Evolutionary biology: volume 6* (eds T Dobzhansky, MK Hecht, WC Steere), pp. 381–398. New York, NY: Springer.
- Dobzhansky T, Hecht MK, Steere WC (eds). 1972 *Evolutionary biology: volume 6*. New York, NY: Springer.
- Ruvolo M, Seielstad M. 2001 'The apportionment of human diversity' 25 years later. In *Thinking about evolution: historical, philosophical, and political perspectives* (eds RS Singh, CB Krimbas, DB Paul, J Beatty), pp. 141–151. Cambridge, UK: Cambridge University Press.
- Brown RA, Armelagos GJ. 2001 Apportionment of racial diversity: a review. *Evol. Anthropol.* **10**, 34–40. (doi:10.1002/1520-6505(2001)10:1<34::aid-ewan1011>3.0.co;2-p)
- Rosenberg NA. 2018 Variance-partitioning and classification in human population genetics. In *Phylogenetic inference, selection theory, and history of science: selected papers of A. W. F. Edwards with commentaries* (ed. RG Winther), pp. 399–404. Cambridge, UK: Cambridge University Press.
- Novembre J. 2022 The background and legacy of Lewontin's apportionment of diversity. *Phil. Trans. R. Soc. B* **377**, 20200406. (doi:10.1098/rstb.2020.0406)
- National Human Genome Research Institute. Activity 1: Genetic Variation in Populations. *National Human Genome Research Institute*. See <https://www.genome.gov/25019961/online-education-kit-activity-1-genetic-variation-in-populations> (accessed on 29 June 2021).
- Herbes-Sommers C, Strain TH, Smith L. 2003 *RACE—The power of an illusion*. Background readings. See https://www.pbs.org/race/000_About/002_04-background-01-04.htm (accessed on 7 September 2021).
- Huckins G. 2021 Richard Lewontin leaves a legacy of fighting racism in science. *Wired*. See <https://www.wired.com/story/richard-lewontin-leaves-a-legacy-of-fighting-racism-in-science/>.
- Williams DR. 1997 Race and health: basic questions, emerging directions. *Ann. Epidemiol.* **7**, 322–333. (doi:10.1016/s1047-2797(97)00051-3)
- Berry JW, Berry JW, Poortinga YH, Segall MH, Dasen PR. 1992 *Cross-cultural psychology: research and applications*. Cambridge, UK: Cambridge University Press.
- Williams DR. 1999 Race, socioeconomic status, and health. The added effects of racism and discrimination. *Ann. N. Y. Acad. Sci.* **896**, 173–188. (doi:10.1111/j.1749-6632.1999.tb08114.x)
- Jost L. 2006 Entropy and diversity. *Oikos* **113**, 363–375. (doi:10.1111/j.2006.0030-1299.14714.x)
- Bornmann L. 2014 Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *J. Informetr.* **8**, 895–903. (doi:10.1016/j.joi.2014.09.005)
- Belter CW. 2015 Bibliometric indicators: opportunities and limits. *J. Med. Libr. Assoc.* **103**, 219–221. (doi:10.3163/1536-5050.103.4.014)
- Ellegaard O, Wallin JA. 2015 The bibliometric analysis of scholarly production: how great is the impact? *Scientometrics* **105**, 1809–1831. (doi:10.1007/s11192-015-1645-z)
- Levins R, Lewontin R. 1985 *The dialectical biologist*. Cambridge, MA: Harvard University Press.
- Jensen CS. 2016 The dark citations of TODS papers and what to do about it. *SIGMOD Rec.* **45**, 69–70. (doi:10.1145/3003665.3003680)
- Gould SJ, Lewontin RC. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598. (doi:10.1098/rspb.1979.0086)
- Lewontin RC. 1974 *The genetic basis of evolutionary change*. New York, NY: Columbia University Press.
- Lewontin RC, Rose SPR, Kamin LJ. 1984 *Not in our genes: biology, ideology, and human nature*. New York, NY: Pantheon Books.
- Nei M, Roychoudhury AK. 1972 Gene differences between Caucasian, Negro, and Japanese populations. *Science* **177**, 434–436. (doi:10.1126/science.177.4047.434)
- Nei M, Roychoudhury AK. 1974 Genic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids. *Am. J. Hum. Genet.* **26**, 421–443.
- Spielman RS, Smouse PE. 1976 Multivariate classification of human populations. I. Allocation of Yanomama Indians to villages. *Am. J. Hum. Genet.* **28**, 317–331.
- Smouse PE, Spielman RS. 1977 How allocation of individuals depends on genetic differences among populations. In *Human genetics: Proc. 5th Int. Congr. Human Genetics* (eds S Armendares, R Lisker), pp. 255–260. Amsterdam, The Netherlands: Excerpta Medica.
- Mitton JB. 1977 Genetic differentiation of races of man as judged by single-locus and multilocus analyses. *Am. Nat.* **111**, 203–212. (doi:10.1086/283155)
- Mitton JB. 1978 Measurement of differentiation: reply to Lewontin, Powell, and Taylor. *Am. Nat.* **112**, 1142–1144. (doi:10.1086/283359)
- Lewontin RC. 1978 Single- and multiple-locus measures of genetic distance between groups. *Am. Nat.* **112**, 1138–1139. (doi:10.1086/283357)
- Latter BDH. 1980 Genetic differences within and between populations of the major human subgroups. *Am. Nat.* **116**, 220–237. (doi:10.1086/283624)
- Neel JV. 1981 The major ethnic groups: diversity in the midst of similarity. *Am. Nat.* **117**, 83–87. (doi:10.1086/283687)
- Smouse PE, Spielman RS, Park MH. 1982 Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. *Am. Nat.* **119**, 445–463. (doi:10.1086/283925)
- Nei M, Roychoudhury AK. 1982 Genetic relationship and evolution of human races. *Evol. Biol.* **14**, 1–59.
- Ryman N, Chakraborty R, Nei M. 1983 Differences in the relative distribution of human gene diversity between electrophoretic and red and white cell antigen loci. *Hum. Hered.* **33**, 93–102. (doi:10.1159/000153357)
- Nei M. 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl Acad. Sci. USA* **70**, 3321–3323. (doi:10.1073/pnas.70.12.3321)
- Wilson EO. 1975 *Sociobiology: the new synthesis*. Cambridge, MA: Harvard University Press.
- Nei M. 1977 F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.* **41**, 225–233. (doi:10.1111/j.1469-1809.1977.tb01918.x)
- Ewens WJ. 1979 *Mathematical population genetics*. New York, NY: Springer.
- Rao CR. 1982 Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.* **21**, 24–43. (doi:10.1016/0040-5809(82)90004-1)
- Gould SJ. 1981 *The mismeasure of man*. New York, NY: Norton.
- Seegerstrale U. 1986 Colleagues in conflict: an 'in vivo' analysis of the sociobiology controversy. *Biol. Philos.* **1**, 53–87. (doi:10.1007/BF00127089)
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994 High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457. (doi:10.1038/368455a0)
- Cavalli-Sforza LL, Cavalli-Sforza L, Menozzi P, Piazza A. 1994 *The history and geography of human genes*. Princeton, NJ: Princeton University Press.
- Lande R. 1996 Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* **76**, 5–13. (doi:10.2307/3545743)
- Williams DR, Yan Yu, Jackson JS, Anderson NB. 1997 Racial differences in physical and mental health: socio-economic status, stress and discrimination. *J. Health Psychol.* **2**, 335–351. (doi:10.1177/135910539700200305)
- Avise JC. 2000 *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press.

46. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002 Genetic structure of human populations. *Science* **298**, 2381–2385. (doi:10.1126/science.1078311)
47. Jost L. 2007 Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439. (doi:10.1890/06-1736.1)
48. Jost L. 2008 G(ST) and its relatives do not measure differentiation. *Mol. Ecol.* **17**, 4015–4026. (doi:10.1111/j.1365-294x.2008.03887.x)
49. Sarasvathy SD. 2008 *Effectuation: elements of entrepreneurial expertise*. Northampton, UK: Edward Elgar Publishing.
50. Haney López IF. 1994 The social construction of race: some observations on illusion, fabrication, and choice. *Harvard Civil Rights-Civil Liberties Law Rev.* **29**, 1–62.
51. Delgado R. 1995 Critical race theory the cutting edge. *philpapers.org*. See <https://philpapers.org/rec/DELCRT-2>.
52. Lewontin RC, Hartl DL. 1991 Population genetics in forensic DNA typing. *Science* **254**, 1745–1750. (doi:10.1126/science.1845040)
53. Feldman MW, Lewontin RC. 2008 Race, ancestry, and medicine. In *Revisiting race in a genomic age* (eds BA Koenig, SS-J Lee, SS Richardson, S Richardson), pp. 89–101. New Brunswick, NJ: Rutgers University Press.
54. Fujimura JH, Bolnick DA, Rajagopalan R, Kaufman JS, Lewontin RC, Duster T, Ossorio P, Marks J. 2014 Clines without classes: how to make sense of human variation. *Sociol. Theor.* **32**, 208–227. (doi:10.1177/0735275114551611)
55. Lewontin RC. 1982 *Human diversity*. New York, NY: Scientific American Library.
56. Lewontin RC. 1991 *Biology as ideology: the doctrine of DNA*. Toronto, Canada: House of Anansi.
57. 1975 'Take the World From Another Point of View'. *NOVA* [TV series episode].
58. Kimura M, Crow JF. 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738. (doi:10.1093/genetics/49.4.725)
59. Nei M. 1972 Genetic distance between populations. *Am. Nat.* **106**, 283–292. (doi:10.1086/282771)
60. Nei M. 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590. (doi:10.1093/genetics/89.3.583)
61. Dean M *et al.* 1994 Polymorphic admixture typing in human ethnic populations. *Am. J. Hum. Genet.* **55**, 788–808.
62. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**, 6531–6535. (doi:10.1093/nar/18.22.6531)
63. Vos P *et al.* 1995 AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414. (doi:10.1093/nar/23.21.4407)
64. Lynch M, Milligan BG. 1994 Analysis of population genetic structure with RAPD markers. *Mol. Ecol.* **3**, 91–99. (doi:10.1111/j.1365-294x.1994.tb00109.x)
65. Excoffier L, Smouse PE, Quattro JM. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491. (doi:10.1093/genetics/131.2.479)
66. Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. 1997 An apportionment of human DNA diversity. *Proc. Natl Acad. Sci. USA* **94**, 4516–4519. (doi:10.1073/pnas.94.9.4516)
67. Peakall R, Smouse PE. 2006 genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295. (doi:10.1111/j.1471-8286.2005.01155.x)
68. Rohlf FJ. 1992 *NTSYS-pc numerical taxonomy and multivariate analysis system*. New York, NY: Applied Biostatistics. See <https://www.worldcat.org/title/ntsys-pc-numerical-taxonomy-and-multivariate-analysis-system/odc/709839509>.
69. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959. (doi:10.1093/genetics/155.2.945)
70. Yeh F, Yang R-C, Boyle T, Ye Z, Mao J, Yang R, Boyle T, Ye Z. 1997 *POPGENE, the user-friendly shareware for population genetic analysis*. Edmonton, Canada: Molecular Biology and Biotechnology Centre, University of Alberta. See <https://sites.ualberta.ca/~fyeh/faqs.html>.
71. Mantel N. 1967 The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.
72. Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, de Campos Telles MP, Rangel TF, Bini LM. 2013 Mantel test in population genetics. *Genet. Mol. Biol.* **36**, 475–485. (doi:10.1590/S1415-4752013000400002)
73. Hamrick JL, Godt MJW. 1990 Allozyme diversity in plant species. In *Plant population genetics, breeding, and genetic variation in plants* (eds ADH Brown, MT Clegg, AL Kahler, BS Weir), pp. 43–63. Sunderland, MA: Sinauer Associates.
74. Nei M, Roychoudhury AK. 1974 Sampling variances of heterozygosity and genetic distance. *Genetics* **76**, 379–390. (doi:10.1093/genetics/76.2.379)
75. Appiah A. 1985 The uncompleted argument: Du Bois and the Illusion of Race. *Crit. Inq.* **12**, 21–37.
76. Fullerton SM. 2007 On the absence of biology in philosophical considerations of race. In *Race and epistemologies of ignorance* (eds S Sullivan, T Tuana), pp. 241–258. Albany, NY: State University of New York Press.
77. National Center for Human Genome Research. 1990 *Understanding our genetic inheritance: The US Human Genome Project, The first five years FY 1991–1995*. OSTI.GOV Technical Report. (doi:10.2172/6958032)
78. Edwards AWF. 2003 Human genetic diversity: Lewontin's fallacy. *Bioessays* **25**, 798–801. (doi:10.1002/bies.10315)
79. Winther RG. 2018 Selection theory. In *Phylogenetic inference, selection theory, and history of science: selected papers of A. W. F. Edwards with commentaries* (ed. RG Winther), pp. 407–427. Cambridge, UK: Cambridge University Press.
80. Angier N. 2000 Do races differ? Not really, genes show. *New York Times*, 22 August. See <https://www.nytimes.com/2000/08/22/science/do-races-differ-not-really-genes-show.html>.
81. Cavalli-Sforza LL, Wilson AC, Cantor CR, Cook-Deegan RM, King MC. 1991 Call for a worldwide survey of human genetic diversity: a vanishing opportunity for the Human Genome Project. *Genomics* **11**, 490–491. (doi:10.1016/0888-7543(91)90169-f)
82. Lewontin RC. 2001 *It ain't necessarily so: the dream of the human genome and other illusions*. New York, NY: New York Review of Books.
83. Priem J, Piwowar HA, Hemminger BM. 2012 Altmetrics in the wild: using social media to explore scholarly impact. *arXiv* 1203.4745 [cs.DL]. (doi:10.48550/arXiv.1203.4745)
84. Attention for 'Chapter 14: The Apportionment of Human Diversity'. *Altmetric*. See <https://www.altmetric.com/details/6453683/chapter/2389105> (accessed on 4 July 2021).
85. Attention for 'Human genetic diversity: Lewontin's fallacy'. *Altmetric*. See <https://www.altmetric.com/details/628085> (accessed on 4 July 2021).
86. Carlson J, Harris K. 2020 Quantifying and contextualizing the impact of bioRxiv preprints through automated social media audience segmentation. *PLoS Biol.* **18**, e3000860. (doi:10.1371/journal.pbio.3000860)
87. Charlesworth B, Charlesworth D. 2017 Population genetics from 1966 to 2016. *Heredity* **118**, 2–9. (doi:10.1038/hdy.2016.55)
88. Pickrell JK, Reich D. 2014 Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet.* **30**, 377–389. (doi:10.1016/j.tig.2014.07.007)
89. Herrnstein RJ, Murray CA. 1994 *The bell curve: intelligence and class structure in American life*. New York, NY: Free Press.
90. 1974 SDS will challenge Prof. Herrnstein to debate with Lewontin on genetics. *The Harvard Crimson*, 1 March. See <https://dev.thecrimson.com/article/1974/3/1/sds-will-challenge-prof-herrnstein-to/>.
91. Block N. 1995 How heritability misleads about race. *Cognition* **56**, 99–128. (doi:10.1016/0010-0277(95)00678-r)
92. 1994 Politics, Race, 'Science'. *The Washington Post*, 1 November. See <https://www.washingtonpost.com/archive/opinions/1994/11/01/politics-race-science/c35c65fc-6ba6-40ad-8289-5914e43af467/>.
93. Murray C. 2005 The inequality taboo. *WSJ Online*. See <https://www.wsj.com/articles/SB122538879063584839> (accessed on 11 September 2021).
94. Du Bois WEB. 1897 *The conservation of races*. Washington, DC: The American Negro Academy Occasional Papers, No. 2.

95. Boas F. 1940 *Race, language, and culture*. New York, NY: Macmillan.
96. Warne RT, Astle MC, Hill JC. 2018 What do undergraduates learn about human intelligence? An analysis of introductory psychology textbooks. *Arch. Sci. Psychol.* **6**, 32–50. (doi:10.1037/arc0000038)
97. Li JZ *et al.* 2008 Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104. (doi:10.1126/science.1153717)
98. Edge MD, Rosenberg NA. 2015 Implications of the apportionment of human genetic diversity for the apportionment of human phenotypic diversity. *Stud. Hist. Philos. Biol. Biomed. Sci.* **52**, 32–45. (doi:10.1016/j.shpsc.2014.12.005)
99. Novembre J, Peter BM. 2016 Recent advances in the study of fine-scale population structure in humans. *Curr. Opin. Genet. Dev.* **41**, 98–105. (doi:10.1016/j.gde.2016.08.007)
100. Biddanda A, Rice DP, Novembre J. 2020 A variant-centric perspective on geographic patterns of human allele frequency variation. *Elife* **9**, e60107. (doi:10.7554/eLife.60107)
101. Roseman CC. 2021 Lewontin did not commit Lewontin's fallacy, his critics do: why racial taxonomy is not useful for the scientific study of human variation. *Bioessays* **43**, e2100204. (doi:10.1002/bies.202100204)
102. Shen H, Feldman M. 2022 Diversity and its causes: Lewontin against racism and biological determinism. *Phil. Trans. R. Soc. B* **377**, 20200417. (doi:10.1098/rstb.2020.0417)
103. Cosentino G. 2020 From Pizzagate to the great replacement: the globalization of conspiracy theories. In *Social media and the post-truth world order: the global dynamics of disinformation* (ed. G Cosentino), pp. 59–86. Cham, Switzerland: Springer International Publishing.
104. Pape RA. 2021 What an analysis of 377 Americans arrested or charged in the Capitol insurrection tells us. *The Washington Post*, 6 April. See <https://www.washingtonpost.com/opinions/2021/04/06/capitol-insurrection-arrests-cpost-analysis/>.
105. Eligon J. 2019 The El Paso Scream, and the racist doctrine behind it. *New York Times*, 7 August. See <https://www.nytimes.com/2019/08/07/us/el-paso-shooting-racism.html>.