Rediscovering the 1%: Knowledge Infrastructures and the Stylized Facts of Inequality[1]

Daniel Hirschman
Brown University

In the 2000s, newly analyzed tax data revealed that top incomes in the United States—summarized as the 1%—had been rising dramatically upward since the early 1980s. Why did it take social scientists two decades to identify this trend and to incorporate it into debates about contemporary income inequality? Drawing on insights from the history and sociology of science, I argue that the social sciences rely on *knowledge infrastructures* to monitor trends and identify stylized facts. These infrastructures collect, process, and distribute data in ways that channel sustained attention to particular problems while rendering other potential observations out of focus. Like other infrastructures, they have significant inertia: initial design choices become locked in and shape the kinds of data readily available to future researchers. Thus economic knowledge infrastructures constructed in the mid-20th century, while identifying some forms of increasing income inequality, were incapable of tracking top incomes, which created the conditions for missing the rise of the 1%.

AN UNEXPECTED FINDING

In 2003, economists Thomas Piketty and Emmanuel Saez published a surprising finding: the top 1% of income earners in the United States were now receiving about 15% of all income, a share almost twice as large as 30 years earlier. Perhaps even more surprisingly, top income earners were now receiving

[1] Funding for this research was provided by the National Science Foundation (grant 1229894), the Rackham Graduate School at the University of Michigan, and the Program in Business, Entrepreneurship, and Organizations at Brown University. Vanessa Garcia and Jime Gonzalez provided excellent research assistance. This project began as part of a dissertation advised by Greta Krippner, John Carson, Gabrielle Hecht, Jason Owen-Smith, and Mark Mizruchi, and I

 https://doi.org/10.1086/718451

their largest share since the Great Depression. Even before the finding appeared in the prestigious *Quarterly Journal of Economics*, the columnist and noted economist Paul Krugman had already written about it in the *New York Times Magazine*.[2] Although scholars had observed an increase in income inequality in the 1980s and 1990s, Krugman argued that Piketty and Saez's research reframed the debate by offering a new characterization of inequality that did not fit easily into existing theories about the role of education, globalization, race, or gender. Over the next decade, the growth of the top 1% spurred new research programs in economics, political science, and sociology. Beyond the academy, "the 1%" entered mainstream political discourse and even helped to frame a new political identity, as the Occupy Wall Street movement declared, "We are the 99%!" (Gould-Wartofsky 2014).

How was it possible that Piketty and Saez's finding was so surprising? The trend of growth in the top 1% of incomes began in the 1980s and was at least potentially visible in publicly available annual IRS tax data. Piketty and Saez did not invent any fancy new statistical techniques but updated an analysis published by Simon Kuznets half a century earlier (see figs. 1 and 2).[3] Indeed, as discussed in detail below, a few economists did identify increases in top incomes in the mid-1990s (e.g., Feenberg and Poterba 1993). Yet that work was isolated and had limited impact on later research, while Piketty and Saez's finding transformed the conversation and was treated as a surprise. Why did economists and other social scientists know so much about other forms of income inequality—median incomes, race and gender gaps, returns to a college degree—but so little about top incomes? What was it about the way scholars studied inequality that made research on top incomes so scarce in the 1980s–90s and thus made it possible for most scholars to miss the rise of the 1%? To answer these specific empirical questions, I develop a framework for addressing a more general one: How do past research priorities shape future knowledge production? I argue that these priorities become institutionalized in *knowledge infrastructures*—"robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge

am grateful for their support. I thank audiences at the University of Michigan, Brown, Harvard University, the University of Notre Dame, and Yale University for helpful comments on previous drafts. Many other readers have provided generous feedback, including the *AJS* reviewers. I thank especially Elizabeth Popp Berman, Jamie Budnick, Marion Fourcade, and Zach Griffen for offering extraordinary advice and feedback over the entire life course of this project.
Direct correspondence to Daniel Hirschman, Department of Sociology, Brown University, Maxcy Hall, Providence, Rhode Island 02912. Email: daniel_hirschman@brown.edu

[2] Krugman (2002), and other commenters, relied on Piketty and Saez's NBER working paper, which was released in 2001.

[3] Throughout this article, I emphasize Piketty and Saez's early work on top *income* inequality rather than Piketty's (2014) later research on *wealth* inequality. See McGoey (2017) for an excellent discussion of the history of that later project with a similar theoretical interest in evidence, absence, and ignorance.

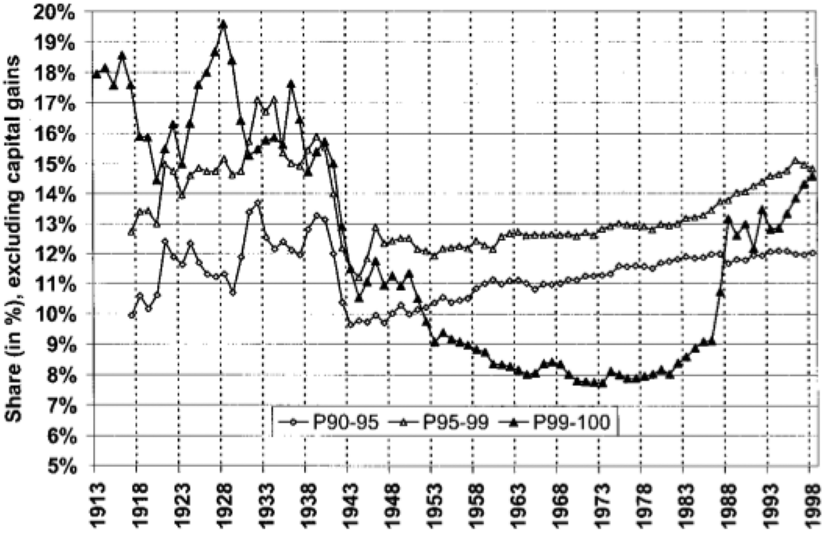

FIG. 1.—Piketty and Saez's (2003, p. 12) famous chart showing the growth of top incomes in the 1980s–90s.

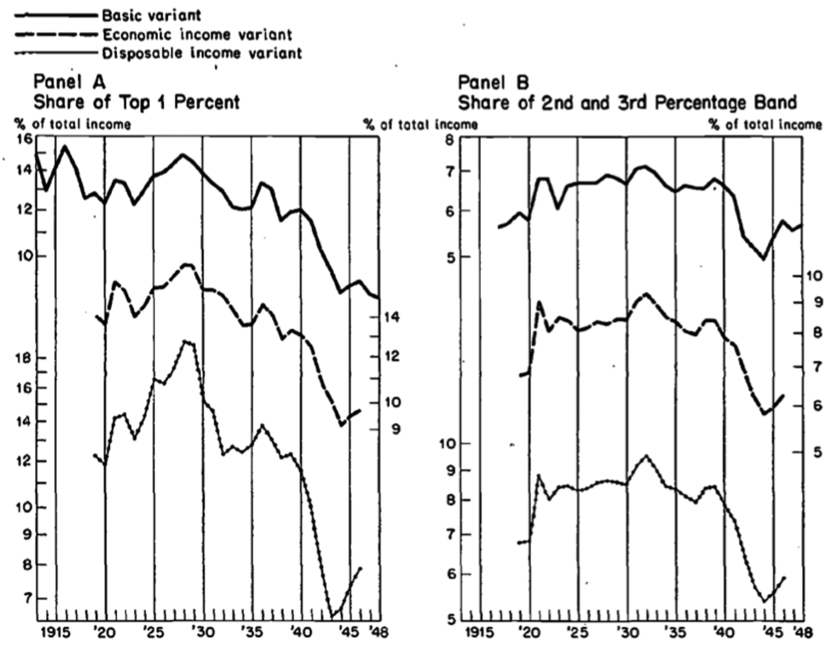

FIG. 2.—Kuznets's (1953, p. 33) less famous chart showing the downward trend of top incomes in the 1930s–40s.

about the human and natural worlds" (Edwards 2010, p. 17)—which enable certain kinds of knowledge production while simultaneously channeling researchers away from questions not readily answerable within the framework of that infrastructure.

Drawing on the history of macroeconomics and labor economics, I show how mid-20th-century economic theory, practicalities of data collection, and economic policy priorities shaped the creation of two knowledge infrastructures that made visible certain aspects of income inequality but were incapable of tracking movements in top incomes. In particular, I show how certain modes of observation that had been common in the 1920s–50s were left out of the prevailing knowledge infrastructures in the 1970s–90s, which in turn rendered top incomes "out of focus" (Peterson 2017) for inequality scholars.

To make sense of this case, I draw on and expand interdisciplinary theorizing about knowledge infrastructures and how they shape the trajectory of scientific fields (Edwards 2010, 2019; Edwards et al. 2013; Ribes and Polk 2015; Boyce 2016). I focus here on knowledge infrastructures built to monitor trends (such as weather stations or national censuses) in contrast to experimental systems typically deployed in laboratories and designed to uncover features of the human and natural worlds that are believed to be stable phenomena (Hacking 1983; Rheinberger 1994). I argue that the social sciences rely heavily on monitoring infrastructures because they make possible the identification of *stylized facts*, empirical regularities that substitute for phenomena (Hirschman 2016*a*).

I offer an account of how knowledge infrastructures collect, process, and distribute data and, in so doing, facilitate specific forms of knowledge production. Not all research is undergirded by an existing knowledge infrastructure, but research that works with and through such infrastructure benefits from the upfront investments made in gathering and cleaning the data, and from the work infrastructures do to facilitate widespread understanding and use of the data. The construction of knowledge infrastructures, like that of physical infrastructures, is a complex process that must resolve competing theoretical, political, and practical demands. Once constructed, knowledge infrastructures acquire significant inertia. In order to monitor trends, this infrastructure must measure the same things over time at the cost of lost flexibility. Researchers whose work (and thus careers and influence) was made possible by a knowledge infrastructure can and do mobilize to secure the resources needed to maintain it, but not necessarily to expand it to meet other needs. As a result, past priorities shape existing knowledge infrastructures that in turn channel researcher attention toward some problems and away from others. Attending to the dynamics of knowledge infrastructures identifies a novel route by which the history of a field is institutionalized in ways that shape future knowledge production.

The remainder of this article proceeds as follows. First, I offer a novel synthetic theorization of the dynamics of knowledge infrastructures. Second, I outline an analytical strategy for approaching the rise of the 1% in inequality debates, arguing that this case can be usefully understood as an example of the emergence of a new stylized fact. Third, I turn to the case itself. I follow the intertwined histories of macroeconomics, labor economics, and the production of official statistics to show how the construction of knowledge infrastructures shaped inequality research. The case is divided into three periods: the 1920s–70s, when both labor economics and macroeconomics take shape and develop specific knowledge infrastructures; the 1980s–2000, when we can see those infrastructures rendering the growth of top incomes out of focus; and 2001–20, when the 1% is rediscovered. In the middle period, I rely on a partial counterexample—a brief moment when top income inequality became a matter of public concern—to theorize the importance of knowledge infrastructures for both observing economic life and for providing sustained attention to particular features. In the discussion, I offer suggestions for how studying knowledge infrastructures can help make sense of other features of the development of inequality knowledge, such as persistently ignored calls for research integrating organizational perspectives.

KNOWLEDGE INFRASTRUCTURES

Historians and sociologists of science refer to systems of observation and measurement as *knowledge infrastructures*. Edwards (2010) focuses on the knowledge infrastructures that make possible claims about global climate through the integration of heterogeneous forms of data, from contemporary observatories making CO_2 measurements to 19th-century naval logs with measurements of ocean surface temperatures. In American social science, central knowledge infrastructures include the decennial census, the National Income and Product Accounts (NIPA), and a handful of long-running surveys like the Current Population Survey, the American National Election Studies, and the Panel Study on Income Dynamics. Identifying these systems as infrastructures highlights several features, in particular that they are "routine, reliable, and widely shared" (Edwards 2017, p. 36). These knowledge infrastructures enable the observation of social and economic life over time, and thus the identification of trends concerning fertility, economic growth, gender inequality, and so on.[4]

[4] Here, I focus on infrastructures designed for monitoring rather than experimental systems. That is, I am interested in the dynamics of knowledge infrastructures built to track the social or natural world "out there" as it changes rather than experimental systems built in laboratories to detect and isolate specific, seemingly immutable, features of the natural (or, less frequently, social or psychological) world (on experimental systems see

To facilitate research, knowledge infrastructures perform three main tasks: they collect, process, and distribute data. At each step of this process, the designers and maintainers of the infrastructure must make consequential choices about what data to collect, how to process it, and how to distribute it. Each step also requires resources, most notably funding, which is typically provided by foundations or the government (either directly for knowledge infrastructures housed inside state agencies or in the form of grants to other organizations). If successful, knowledge infrastructures facilitate knowledge claims by lending legitimacy, cultivating familiarity, and thus increasing the "doability" (Fujimura 1987) of research that builds on their capacities.

At the center of a knowledge infrastructure are its mechanisms for collecting and recording information. These mechanisms arise from choices made by the architects of the knowledge infrastructure, who are informed by research agendas, professional values, and political priorities. Contemporaneous theoretical and empirical debates may weigh especially heavily on design choices. These choices are about the basic ontology of the infrastructure—the categories studied, the units of analysis—as well as the scope and scale (global, national, local, etc.), and emphasis or focus (Peterson 2017). All of these choices are constrained by the practical difficulties of data collection. For example, Bowker and Star (1999, p. 46) note that the original International Classification of Diseases was limited to "200 diseases not because of the nature of the human body and its problems but because this was the maximum number that would fit on the large census sheets then in use."

Knowledge infrastructures are typically built around a "kernel," a relatively stable center to the infrastructure that embeds core parts of the initial research agenda (Ribes and Polk 2015). The design of the "kernel renders certain objects easily inspected; others are a greater challenge, and still others cannot be inspected at all." (Ribes and Polk 2015, p. 236; see also Peterson 2017). For example, the Panel Study on Income Dynamics (PSID) included a larger-than-representative sample of Black households in its kernel (the initial panel of respondents) because the study emerged from the 1960s War on Poverty and oversampled poor, nonwhite households (McGonagle et al. 2012). Thus the PSID was better suited to tracking trends in Black-White disparities than many other national surveys.

Kernels are not completely fixed; they may be extended or elaborated, but doing so may be costly, and the infrastructure is not infinitely flexible (Ribes and Polk 2015). In the case of the PSID, attempts to extend this kernel to better track other groups have met with mixed success. For example, in 1990, the PSID added a new sample of households with heads from Mexico,

---

Hacking [1983], Rheinberger [1994]). While the two kinds of systems share many features (such as the need to make choices about collecting and processing data), monitoring infrastructures face specific additional pressures toward inertia that are theoretically important for understanding how these infrastructures shape the trajectory of research.

Cuba, and Puerto Rico, but this subsample was expensive to construct and maintain, and was dropped in 1995. Many other national surveys have similar limitations rooted in their existing designs (Clemens and Garvey 2020).

Because knowledge infrastructures require substantial investments to build and maintain, academics typically must also secure buy-in from private foundations or the state and thus must justify their research according to the logics of those actors. The National Institutes of Health, for example, although long a funder of large-scale public health surveys, famously refused to fund research on topics perceived to be controversial, like sexuality.[5] Alternatively, academics may build "second-order" knowledge infrastructures (Boyce 2016) which rely on "first-order" data collected for other purposes (thus requiring fewer new resources), but in so doing they inherit whatever priorities and limitations are present in those first-order infrastructures. For example, before World War II, scholars studying the personal distribution of income using tax data were limited in their ability to see the whole distribution as the majority of Americans did not pay income taxes.

After collecting data, knowledge infrastructures typically process the data, producing "clean" data sets (Plantin 2019). This processing may meet multiple objectives, from improving data quality by weeding out errors in the data collection process and interpolating missing values (Bokulich 2020), to protecting respondent privacy by removing identifying information and collapsing small cells, to simplifying the data to make it more user-friendly by recoding response categories (Budnick 2020). These choices may make the data more useful for some purposes, but they simultaneously require discarding or filtering out information and thus necessarily limit certain avenues of research (Bowker and Star 1999). Downstream researchers may be unaware of the choices made in processes of cleaning and thus unaware of how such practices may affect their results or interpretations (Borgman et al. 2014). In the context of research on income inequality, as I discuss more below, surveys frequently "top code" high incomes, reporting such incomes as "above X" rather than reporting the exact value, which renders changes in the top of the income distribution invisible.

Finally, after collecting and processing the data, knowledge infrastructures also make that data available and usable to the broader community. In addition to sharing the data (itself not a trivial task, especially before the advent of the internet), knowledge infrastructures also typically provide various forms of metadata, such as codebooks, questionnaires, and other documentation needed to facilitate the data's use (Downey, Eschenfelder, and Shankar 2019). Some knowledge infrastructures do much more, offering

[5] A notable example of this refusal is the revocation of funding of what eventually became the National Health and Social Life Survey (Laumann, Michael, and Gagnon 1994; Budnick 2020).

dedicated classes on how to use and interpret the data and thus cultivating the research community's expertise and familiarity with the data (Shankar, Eschenfelder, and Downey 2016). For example, in addition to teaching classes on a range of generic statistical methods, the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan also currently offers regular workshops to train users of the PSID. Funding agencies like the National Science Foundation often provide resources specifically to support such activities as part of large grants for collecting or maintaining large-scale survey data (Solovey 2020). These activities build a community of users who in turn increase the usefulness of the data to a given researcher. Other scholars already understand what the data can do, they may share standardized code to construct new variables or aid analysis of interest to researchers, and they are qualified to review papers that make use of it, all of which reduces barriers to using such data while increasing its legitimacy. In turn, the increased use of the data justifies continued funding support, as advocates for the knowledge infrastructure can point to the growing list of papers as evidence of influence and importance.

At each stage of this process—collection, processing, and distribution—architects and maintainers of knowledge infrastructures make choices that are potentially consequential for the kinds of knowledge produced by a research community. I note four main consequences of this work.

First, knowledge infrastructures are productive of particular forms of knowledge. They enable research communities to produce knowledge more easily, to achieve greater consensus, and especially to identify and monitor trends over time. The capacity to monitor trends is particularly important because of how the social sciences rely on identifying and theorizing stylized facts (Hirschman 2016*b*). Stylized facts take a variety of forms, including correlations, rates of incidence, or trends. Since the 1960s, economists (and some sociologists and political scientists) have explicitly discussed their theoretical models in terms of the stylized facts they are designed to explain. Prominent economic stylized facts include labor's relatively constant share of national income (one of the original stylized facts identified by Kaldor [1961], who coined the term), the increasing returns to a college degree, the gender pay gap, the racial wealth gap, and, as I will argue below, the rise of the 1%. These facts are understood as stable enough to merit theorizing, but also mutable enough to merit systems of observations and measurement designed to track their stability.

Second, the capacity of a knowledge infrastructure to measure change over time entails a certain level of stability: to measure a trend, the same features of the social or natural world must be measured in the same way over the period of interest. Knowledge infrastructures make this possible. But this need for stability simultaneous generates "tyrannies of inertia" (Bowker and Star 1999, p. 50). Even as researchers recognize limits in existing knowledge infrastructure, perhaps related to novel empirical or theoretical questions, that infrastructure is difficult and expensive to reorient, and doing so may break the perceived continuity of the measures. National income accountants point to these tensions when they justify continuing to omit estimates of the market value of unpaid housework from Gross Domestic Product (GDP) calculations (DeRock 2021). This tension is particularly heightened in domains where the terrain itself is shifting rapidly. Researchers tracking attitudinal changes over time rely on surveys asking identical questions over many years, yet the wording of questions may itself become outmoded. For example, Bonilla-Silva (2017) argues that questions designed to measure racial attitudes in the 1960s no longer accurately capture contemporary forms of racism, and Budnick (2020) tracks the persistence of outdated categories in questions about attitudes about nonheterosexuality. Much as political scientists and sociologists have identified a process of policy feedback, whereby constituents mobilize to defend policies that have benefited them (Campbell 2012), researchers whose work relies on a particular knowledge infrastructure become a key constituency that mobilizes to secure support to maintain the infrastructure (Solovey 2020); they may even resist expanding it or supporting new infrastructures that they feel might compete with or distract from their core mission (DeRock 2021). As a result of this strong tendency toward inertia combined with changes in the social world itself, knowledge infrastructures may, over time, experience a growing "concept-measurement gap" (Mügge and Linsi 2021).

Third, this inertia implies that the choices made at the initial construction of a knowledge infrastructure may have especially enduring consequences. The theoretical, empirical, and political interests that motivate the construction of the knowledge infrastructure, and the political bargains required to secure the resources to establish the infrastructure, are worked into the decisions about data collection, processing, and distribution. These "quiet victories of infrastructure builders inscribing their politics into the systems" (Bowker and Star 1999, p. 50) occur in part through settling debates over processes of classification required to build the infrastructure. Similarly, choices made about the kernel are very difficult to change. As a result, old theories and priorities may structure present understandings even if the research community no longer accepts those theories. For example, as I discuss below, 20th-century NIPA statistics were developed and standardized along explicitly Keynesian lines, and the choices made in this process structured debates in macroeconomics even after Keynes's ideas were rejected or amended.

Fourth, and finally, knowledge infrastructures produce ignorance as well as knowledge. Because knowledge infrastructures facilitate some forms of research, enhancing their doability (Fujimura 1987) and monopolizing resources and interest, other research remains "undone" (Frickel et al. 2010). This process creates ignorance defined as a "domain-based absence of knowledge" (Frickel and Edwards 2014, p. 215). Research priorities at the time an infrastructure is constructed shape what data are *not* collected; in turn, the absence of data about particular phenomena may lead researchers to avoid studying it, which then reinforces the lack of theoretical interest in the topic, and justifies not spending scarce resources to modify the infrastructure to collect that data. In short, knowledge infrastructures are prone to ignorance loops, "iterative cycle[s] of non-knowledge" (Durant 2020, p. 1).

In the following sections, I showcase how these dynamics make sense of the initial empirical puzzle about the rediscovery of the 1%. I then use insights from that case to suggest how attending to knowledge infrastructures can make sense of other patterns in the history of inequality knowledge.

ANALYTICAL STRATEGY

To motivate this analysis, I first establish that there has been a surge of scholarly and public interest in top incomes since 2001. Here, I draw on an original hand-coded data set of newspaper articles about income inequality along with academic journal citation data. This analysis establishes that scholars and public discourse detected increased inequality in the mid-1980s but did not attend to increasing top incomes until the 2000s.

Having established this transformation in inequality discourse, I then go back to the early 20th century to trace the development of the knowledge infrastructures operating in the 1980s–90s. I trace the histories of U.S. labor economics, macroeconomics, and the production of official economic statistics over three periods (1920s–70s, 1980s–2000, and 2001–20). I draw on a variety of sources in addition to popular and scholarly articles, including conference proceedings, edited volumes, publications from the Department of Commerce, transcripts of congressional and archival sources collected as part of a larger project on the history of macroeconomics (see Hirschman 2016*a* for details). I use these documents to show how, in the first period, the acceptance of particular economic theories (primarily Keynesian approaches to macroeconomics and human capital theory in labor economics) alongside political imperatives (the War on Poverty) shaped the kinds of data that economists collected and analyzed (national income accounts and household surveys, respectively), and the stylized facts that economists identified. Other kinds of data were not entirely excluded, but they were less clearly connected to major theoretical debates and thus capable of being ignored. In the second period, I add to these sources an analysis of public debate around income inequality, and in particular, an important event in 1992 where top incomes briefly became a subject of political contention before receding into the background. Finally, I move to the most recent period and ask what we can learn from the rediscovery of the 1% through the reception of Piketty and Saez's work. This section showcases the ongoing efforts to solidify the knowledge infrastructure developed by Piketty and Saez to measure top incomes as well as the development of new academic research programs and new social movements that emphasized top incomes as a central facet of income inequality.

ESTABLISHING THE PHENOMENON

Before explaining *how* we were collectively capable of being surprised by the rise of top incomes, I first document *that* we were surprised. In this section, I draw on a hand-coded analysis of income inequality discourse and citation data from economics to establish the following empirical claims:

1. Scholars and the media recognized that income inequality was increasing in the 1980s–90s.
2. Scholars and the media did not recognize the particular growth of top incomes in the 1980s–90s.
3. In the 2000s, economists and other social scientists increasingly recognized the growth of top incomes and began to theorize both its causes and consequences.
4. In 2010–13, public discussions of inequality increasingly recognized the growth of top incomes. In particular, discussions of inequality began to center on the rise of "the 1%."

For claims 1, 2, and 4, I rely on an analysis of inequality discourse from three national newspapers: the *New York Times*, the *Washington Post*, and the *Wall Street Journal*. Working with two research assistants, I constructed and analyzed a corpus of 1,960 articles published between 1980 and 2013.[6] This corpus includes every article in each publication that contained the phrase "income inequality" or "income distribution" anywhere in the text of the article. I use this corpus to ask how inequality was discussed within the set of articles explicitly discussing the concept. This analysis thus complements McCall's (2013) analysis of news magazines, which looks at inequality-related conversations in a broad sample of articles about economic topics and will be discussed further below.

I draw on these data to show how newspapers discussed inequality in the United States as an explicit topic. Of the 1,960 articles identified, 405 were eliminated from the sample because they discussed inequality only outside the United States ($n = 316$) or because they were otherwise deemed irrelevant ($n = 89$)—for example because they used the phrase "income distribution" in the context of companies determining stock dividends rather

[6] The corpus here includes editorials and book reviews as well as traditional news articles. For simplicity, I refer to all of these as "articles."

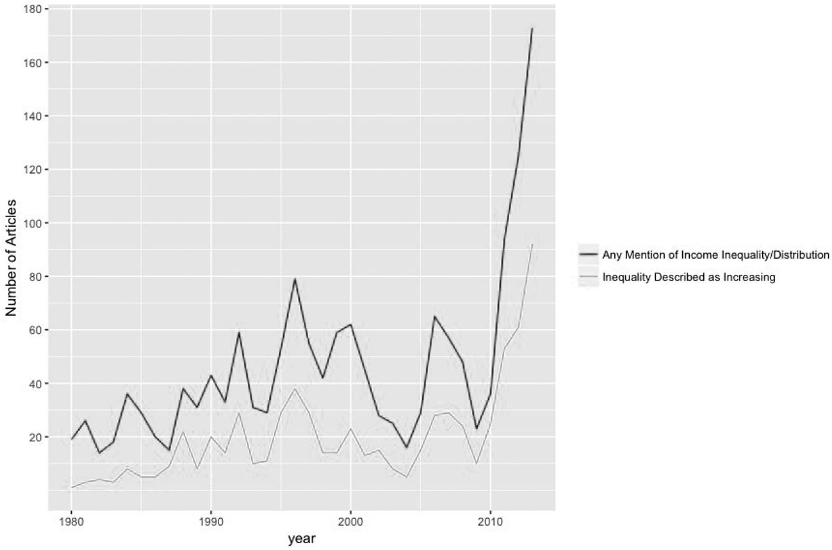

FIG. 3.—Counts of newspaper articles mentioning income inequality and those that specifically describe inequality as increasing.

than in reference to the personal distribution of incomes. The remaining 1,555 articles were coded for how they discussed income inequality, such as whether the article included a claim that inequality was increasing, stable, decreasing, or without a trend, what kinds of inequality were mentioned (gender, race, top incomes), and what specific measures of the income distribution were invoked (median income, top quintile, top 1%).[7] In what follows, I focus on three trends: the overall prevalence of inequality discourse, the recognition that inequality was increasing, and the timing of discussions of the 1% in comparison to other measures of inequality.

Figure 3 displays the distribution of the sample over time, that is, the number of articles mentioning "income inequality" or "income distribution" between 1980 and 2013, and the number of those articles that describe inequality as increasing. This figure shows a dramatic rise in explicit discussions of income inequality in 2010–13, but it also shows substantial discussion in early years, with articles appearing approximately once per week in the 1990s. Figure 3 also shows that, in the early 1980s, most articles about inequality did not identify inequality as increasing. That is, measures of income distribution were discussed without an associated claim that inequality was increasing, stable, or decreasing. Of the 19 articles in the corpus in 1980, just one includes a claim that inequality is increasing, while 18 do not mention a trend. By the

[7] The coded data set and codebook are available on request.

mid-1980s, articles begin to routinely identify inequality as increasing. For example, a 1986 *NYT* story titled “Warnings over Widening Income Gaps” cites a Congressional Research Service study highlighting the declining share of income received by the middle 60% of income earners and the increasing share of income received by the top 20%. Articles including a claim that inequality was stable or decreasing, in contrast, are virtually absent throughout the sample period. Thus figure 3 shows that increasing income inequality was recognized in the mid-to-late 1980s.

In contrast, figure 4 looks at how inequality was discussed in the corpus. This chart shows the number of articles that include a measure of the income of the top quintile, a measure of the incomes of the top 1%, or a mention of the top 1% (with or without a particular measure of their income). Before the publication of Piketty and Saez’s (2001) work, the 1% is mentioned less often than top quintile measures (the main measure of higher incomes, when any are mentioned), and substantially less in most years (including many years with little or no mention). Figure 4 shows how reference to “the 1%” takes off as a framing in the second half of the 2000s and how, after 2010, the term “the 1%” comes to be used on its own, when previously it was almost always associated with a particular measure of income inequality. In the 1980s–90s, the 1% is not yet a a stylized fact about changes in inequality nor an object of explicit political claims making. By 2011, it is. In Hallett, Stapleton, and Souder’s (2019) terminology, “the 1%” becomes a public idea, discussed both as an object of interest (e.g., in debates about exactly how

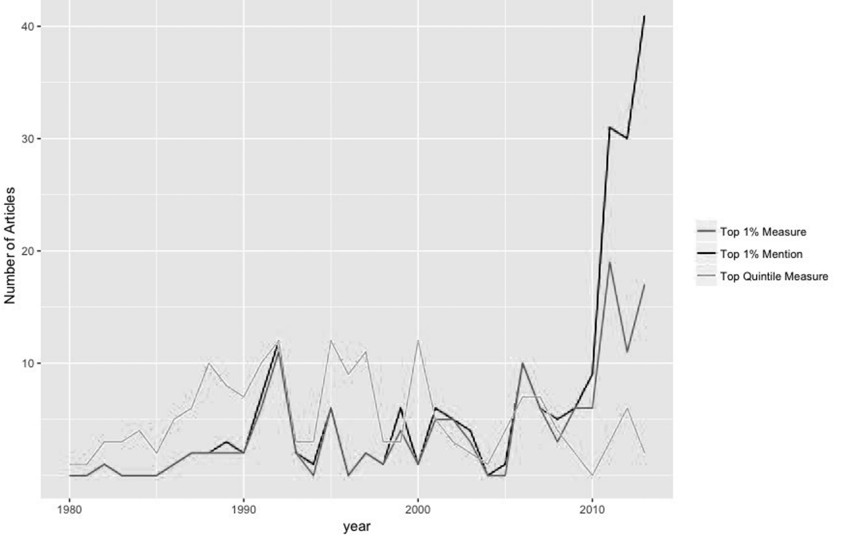

FIG. 4.—Counts of newspaper articles mentioning top incomes.

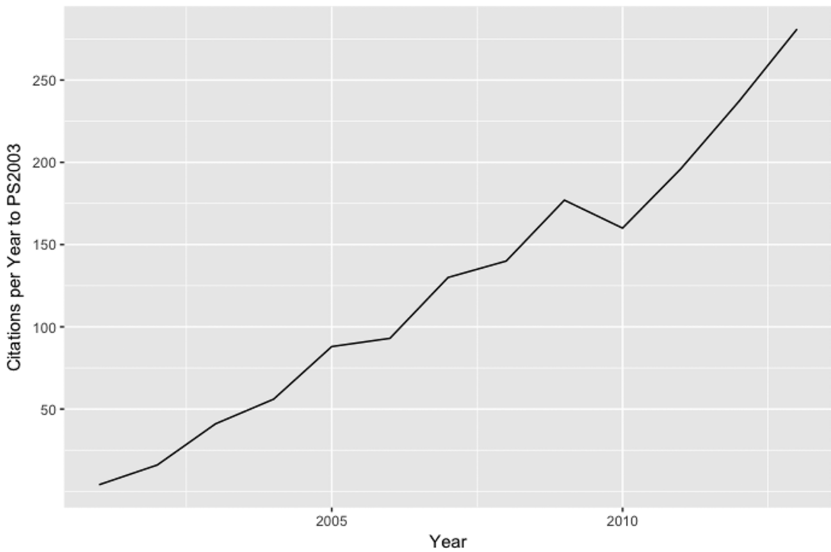

FIG. 5.—Citations per year to Piketty and Saez (2003) in Google Scholar.

much top incomes have grown) and an interpretant (a tool for making sense of other events, such as the fallout of the financial crisis).[8]

Together, figures 3 and 4 show that public discourse about inequality recognized increasing inequality in the mid-1980s, but that discussions of top incomes were minimal until the 2000s and that discussions of the 1% accelerate in the mid-2000s and especially post-2010. To showcase how *academic* discourse changed, figure 5 displays annual citations to Piketty and Saez's 2003 article (including citations in 2001 and 2002 to the working paper version of the article) from Google Scholar. These figures show the rapid uptake of Piketty and Saez's central descriptive finding that top incomes had risen dramatically. Analysis using Web of Science data (which captures a smaller set of citing documents but with higher quality metadata) shows that Piketty and Saez's work has been influential across fields. Of the 1,217 citations tracked there, 46% were from publications in economics, 9% from sociology, 8% from political science, and the rest from an assortment of other fields. Later, I offer a more detailed discussion of how this finding influenced research programs in political science, sociology, and especially economics.

The analysis shared in figures 3–5 establishes the puzzle to be explained: despite recognizing increasing income inequality since the mid-1980s, scholars

[8] A partial exception to this chronology appears in 1992, when discourse about top incomes briefly spikes before receding. This exception, related to a particular debate in the 1992 presidential election, is discussed extensively below.

and the media did not see the rise of the 1% until the 2000s. Why not? To answer this question, I return to the early 20th century to trace how economists developed the knowledge infrastructures that were in place in the 1980s.

BUILDING KNOWLEDGE INFRASTRUCTURES

Before the 1920s, the United States produced none of the aggregate economic statistics we now take for granted. Data—official or privately produced—on employment, inflation, income distribution, and growth were largely unavailable throughout the 1800s, and only became available at the state or local level in the 1890s–1910s (Stapleford 2009; Card 2011). In the 1920s–40s, the United States government began to produce consistent series of national data on inflation (the Consumer Price Index), unemployment (via the Current Population Survey, or CPS), and economic growth (in the National Income and Product Accounts).

In 1920, a group of economists founded the National Bureau of Economic Research (NBER) with the explicit goal of producing timely, objective data about the economic system for use by policymakers, businesses, and academics alike (Rutherford 2011). NBER's first study (NBER 1921) used newly available income tax data to examine both national income *and* income distribution. Somewhat ironically in retrospect, this study focused heavily on top income earners precisely because only top incomes were well-covered by tax data (see fig. 6 for an example of their findings).[9] In the 1930s, work on national income was transferred from NBER to the Department of Commerce, and distributional issues would take a backseat to better, timelier measurement of total national income.

National, survey-based statistics began in earnest with efforts to measure prices and expenditures during World War I (the beginnings of the Consumer Expenditure Survey, used to calculate the Consumer Price Index; see Stapleford 2009). But these efforts but did not become widespread until the 1940s. Survey-based measurement of employment and unemployment, for example, began late in the Great Depression; for most of the 1930s, there were no nationally representative data on unemployment. In the postwar period, the unemployment survey conducted by the Bureau of Labor Statistics was transformed into the modern CPS, which also began to collect income data.

In the same period that official statistical productions took off, the academic field of economics was itself transformed (Fourcade 2009). Just as

[9] In the early 1910s, approximately the top 4% of earners made more than $2,000 per year and thus paid income tax (NBER 1921, p. 113). The income tax did not become a "mass tax" until World War II (Jones 1988).

A CONJECTURAL ESTIMATE OF THE PERCENTAGE OF THE NATIONAL INCOME RECEIVED BY THE HIGHEST FIVE PER CENT. OF INCOME RECEIVERS

1913-1919

Year	Income of the Highest 5% of Income Receivers (Billion Dollars)	Total Individual Income (excluding Corporate Surplus) (Billion Dollars)	Per Cent. of Total Income Received by Highest 5% of Income Receivers
Including Farmers			
1913	$10.6	$32.3	33
1914	10.3	32.0	32
1915	11.1	34.3	32
1916	14.3	41.6	34
1917	14.7	50.5	29
1918	15.4	60.0	26
1919	15.5	64.7	24
Excluding Farmers			
1913	$ 9.9	$28.1	35
1914	9.6	27.8	34
1915	10.4	29.6	35
1916	12.8	35.8	36
1917	13.6	41.7	32
1918	13.9	49.5	28
1919	14.4	53.8	27

FIG. 6.—A 1921 estimate of the income received by the top 5% of income earners based on tax data (NBER 1921, p. 116).

there was little economic data before the 1900s, there was also little "empirical" work in economics, at least in the sense of analysis of large-scale data sets. Inspired by the German Historical School, the American institutionalists began to produce and analyze such data in the 1900s–1920s. Wesley Mitchell was both a central figure in the institutional movement and the first research director of the NBER. Institutionalists would play an important role in integrating private and public data collection efforts, in part through the influence of Herbert Hoover, who advocated such efforts strongly as both secretary of commerce and president as a means for government to coordinate economic activity without the need for regulation or direct control (Alchon 1985; Barber 1985; Rutherford 2011). The institutionalists' direct influence was relatively short-lived, however. The modern divide of macroeconomics and microeconomics emerged in the 1930s, alongside the terms themselves. Institutional economics faded away in the immediate postwar era, replaced by the synthesis of increasingly mathematical neoclassical microeconomics and Keynesian macroeconomics (Yonay 1998; Rutherford 2011). While the institutionalists had been heavily involved in the initial creation of official economic statistics, it would be the Keynesians and neoclassicals who would take charge of finalizing the knowledge infrastructures built around collecting and analyzing those data.

In spite of never acquiring a single iconic statistical representation analogous to GDP for growth, inequality did receive attention from macro and labor economists, who built tools for its observations into their knowledge infrastructures. Macroeconomists focused on the traditional question of *factor shares* and drew on the national accounts to track labor's and capital's relative shares of the total economic pie. Personal income distribution was outside the scope of most macroeconomic research. Labor economists, in turn, relied on survey-based measures to track "gaps" between different kinds of workers, with a special focus on the gaps between college-educated and less educated workers. Income inequality was thus an important topic for labor economists, but they understood it primarily through the lens of the differential productivity of different kinds of workers. Although tax data had formed the basis for early studies of income distribution, by the 1960s, both macroeconomists and labor economists had built infrastructures that relied on other data to observe inequality. The Bureau of Economic Analysis (BEA)[10] did produce tax-data-based statistics on income distribution, but these statistics were not capable of answering the questions economists were interested in, and eventually the statistics were defunded. Thus, by the 1970s, the two dominant knowledge infrastructures for studying income inequality were practically incapable of seeing changes in the top of the personal income distribution.

Factor Shares in Macroeconomics

Macroeconomics, as such, did not exist before the 1930s. Macroeconomics emerged in the 1930s and 1940s at the intersection of monetary theory and business cycle theory, which were joined by a novel interest in economic growth (De Vroey 2016). Here, I focus particularly on how macroeconomics came to largely ignore the personal distribution of income in favor of continued analysis of labor's share of national income.

The distribution of income between the three great classes was arguably the central question of classical political economy (Sandmo 2014). Starting with the writings of the physiocrats in the mid-18th century, political economists divided the economic system into roughly three groups: landlords (who earned rents), employers (who earned profits), and laborers (who earned

[10] For much of this period, the BEA was known as the Office of Business Economics, and before that the Bureau of Foreign and Domestic Commerce. For simplicity, I refer to it as the BEA throughout.

wages). Quesnay's famous 18th century *tableau economique* modeled the flow of goods (or money) from landowners to agricultural workers to artisans and merchants. Although Adam Smith focused less sharply on questions of distribution, David Ricardo (1817, pp. 2–3), went so far as to claim that "in different stages of society, the proportions of the whole produce of the earth which will be allotted to each of these classes, under the names of rent, profit, and wages, will be essentially different. . . . To determine the laws which regulate this distribution, is the principal problem in Political Economy" (see also McNulty 1980, p. 64). Twentieth-century macroeconomics was not nearly so focused on distributional questions as Ricardo or his 19th-century successors, but when macroeconomists looked at inequality they did so through a Ricardian lens.

Although Ricardo and his followers theorized about the distribution of income between the classes—what we now call "factor shares" or "the functional distribution of income" as opposed to the personal distribution of income—they did so without much recourse to economic statistics. Estimating the distribution of income between the great classes became a primary motivation for the construction of national income statistics. This motivation is visible in the first national income statistics published by the Department of Commerce in 1934. These estimates were produced in response to a 1932 U.S. Senate resolution authored by Senator Robert M. LaFollette Jr. of Wisconsin.[11] LaFollette was an outspoken proponent of the need for better economic measurement in general. In an interview given in March 1932, LaFollette connected his push for better measurement of national income to debates about wage cuts in the Great Depression:

> Likewise, with all the talk we have heard from bankers and others about the need for cutting wages and with all the actual wage-cutting that has taken place, we are woefully lacking in any adequate wage statistics. Also, while we are discussing the wages of labor, it is startling that we have no accurate information on the wages which capital is taking in the form of net profits from the point of view of industry as a whole. Furthermore, do you know that we have never had any official estimate of the total national income of the United States and the only authoritative information we have to go on is the estimate of an unofficial agency in 1929?[12]

LaFollette's language echoes Ricardo; the "wages of labor" and the profits of capital are at issue, not the wages of individual laborers or the incomes of particular capitalists. This focus on factor shares was reflected in the presentation of the 1934 report to the Senate, usually characterized as the first official

[11] Senate Resolution 220, 72nd Congress.

[12] "Radio interview between Mr. Charles Ross and Senator Robert M. LaFollette, Jr., over N.B.C., March 15, 1932, 8 p.m." LaFollette Family Collection, Box I:C557, Library of Congress.

publication of national income statistics in the United States.[13] To estimate the sum total of income payments, Simon Kuznets, who led the team that produced the estimates separated incomes first into "labor incomes" (wages), "property incomes" (rents), and "entrepreneurial incomes" (profits) before dividing them into finer subcategories (Kuznets 1934, p. 2). The report's first chart breaks down total national income by type of payment—rents, interest, dividends, salaries, wages—operationalizing the three primary divisions of income into slightly more tractable subdivisions (Kuznets 1934, p. 15). Although the 1934 report did not attempt to estimate the personal distribution of income, Kuznets noted the importance of such estimates: "Economic welfare cannot be adequately measured, unless the personal distribution of income is known" (Kuznets 1934, p. 6). Kuznets's later work, discussed below, made great strides in estimating that personal distribution, but such efforts were not integrated into the production or analysis of national income statistics.

Between the 1934 estimates and the end of World War II, national income statistics took on a decidedly Keynesian character. In part due to the influence of Keynes himself, who oversaw the production of the first official national income statistics in the United Kingdom in the early 1940s, national income statistics became increasingly interwoven with macroeconomic theory (Tily 2009). Rather than emphasizing factor shares of *income*, the national income accounts would increasingly focus on consumption, investment, and government *production*, leading to the famous formula $Y = C + I + G(+NX)$ (gross national product equals consumption plus investment plus government spending [plus net exports]). This formulation also fit with the novel political uses of national income statistics as devices for managing the wartime economy (Lacey 2011) and, later, for tracking economic growth (Collins 2000), rather than as devices for seeing inequality of any sort.

To the extent that postwar macroeconomists, and national income statisticians, did care about the distribution of income, they continued to focus on the role of factor shares. For example, macroeconomist Nicholas Kaldor (1961) identified the stability of labor's share of national income as one of six "stylized facts" about macroeconomics that later models would attempt to explain. The influential research program of Cobb-Douglas regressions looked at the relationship, in the aggregate, between labor's productivity and labor's share of income (McNulty 1980, p. 180; Biddle 2012). Similarly, the dominant macroeconomic forecasting models of the 1950s and 1960s included an analysis of labor's share of national income, but did not include measures of the distribution of personal income (Metcalf 1969).[14] In this sense, the knowledge

[13] This framing ignores an uninfluential and often-forgotten 1926 report by the Federal Trade Commission; see Carson (1971) for details.

[14] Metcalf's own work represents a partial exception; his dissertation attempted to integrate personal distribution of income into standard forecasting models. As Metcalf (1969,

infrastructure and theoretical priorities reinforced each other over time: econometricians and macroeconomists theorized factor shares, which were conveniently already measured in the national accounts, which in turn had been shaped by macroeconomic theorists.

The path from the rise of Keynesianism to the exclusion of personal income data from macroeconomics was not entirely straightforward. One of the central contributions of Keynes's (1936) *General Theory of Employment, Interest, and Money* was the concept of the "marginal propensity to consume." Keynes argued that many dynamics, including the "multiplier" that related increases in government spending to increases in total output, depended on how much consumers would change their spending in response to receiving an additional dollar of income. If the marginal propensity to consume were high, then a bit of extra income created by government spending would multiply many times. Conversely, if the propensity were low, then extra government spending would do little to stimulate the economy. From the very beginning, macroeconomists theorized that the marginal propensity to consume was a function of personal income—individuals who made a lot of money were more likely to save additional income, while poorer individuals would likely spend almost all of it. Such theorizing suggested that the personal distribution of income might be relevant to Keynesian models, and there were some investigations along those lines (e.g., Stone and Stone 1938; see Fixler and Johnson [2012, p. 26] for details). These early efforts revealed small differences between models including and excluding the distribution of income, and macroeconomists largely abandoned the study of the relationship between personal income distribution and the multiplier (Thomas 1992).[15]

Gaps in Labor Economics

Like macroeconomics, labor economics emerged as a recognizable subfield in the first half of the 20th century. As McNulty (1980, p. 2) notes, in the early 19th century, less than 15% of American income earners were employees; by 1980, 90% of income earners were employees. The field of labor economics thus grew as the phenomenon of wage labor itself became dominant. When a recognizable subfield of labor economics did emerge in the late 19th and early 20th centuries, it remained outside of the mainstream. Labor markets

p. 657) notes, "The Brookings model, for instance, limits its consideration of the income distribution to a determination of factor shares. Issues relating directly to how families or income units are distributed by size of income were never raised."

[15] Fixler and Johnson (2012) present a recent exception that ultimately supports this claim; for prior studies on the effect of the personal distribution of income on the marginal propensity to consume, they cite only a handful of papers from the 1950s. See also Auclert and Rognlie (2020).

were seen as something other than the idealized competitive market, dominated by regulations and unions (McNulty 1980). Unsurprisingly, institutionalist economists studied labor markets extensively in the first half of the 20th century, while neoclassical economists largely left the field unexamined. In the 1950s, as the institutionalist movement faded away, a new generation of neoclassical economists turned their attentions to the problems of labor.

Neoclassical economists turned the determination of individual wages into a central question of labor economics. Why did some people earn more money in the labor market than others? The answer developed in the 1960s was "human capital," differential productivity resulting from formal education and on-the-job training. Human capital theories displaced older approaches that either treated the personal distribution of income as resulting from largely random forces, or from inherited ability (Blaug 1976; Teixeira 2007; Sandmo 2014). Here, I trace the history of human capital theory and how it came to study incomes through large-scale surveys focused on measuring wages, work experience, and education. This history points to a deep irony: in the 1920s–50s, research on the distribution of income focused on top incomes *because* there was good tax data that measured them. In the 1960s–90s, studies of income distribution *ignored* top incomes because there was no good survey data capable of measuring them.

To understand the rise of human capital theory, we must first trace two earlier developments: the first statistical studies of the personal distribution of income and the development of the marginal productivity theory of distribution. Although the personal distribution of income was not a central topic to economics in the 19th and early 20th centuries, a small body of statistical work on the subject accumulated. In 1897, Vilfredo Pareto identified a statistical pattern in the distribution of incomes (the Pareto distribution), a particular functional form that seemed to match the shape of the income distribution and in particular its heavy right tail (see discussion in Champernowne [1952]). Later works attempted to estimate the parameters for that distribution, and others, which best fit the observed distribution of income. For example, Champernowne (1952, p. 598) fit a modified Pareto distribution to distribution of income data from the United States in 1918, while Rutherford (1955) fit a probit distribution to the log of income. To the extent that personal incomes seemed to follow particular statistical patterns, economists sought to explain those patterns with random or "stochastic" models. For example, Champernowne (1973) and Wold and Whittle (1957) both develop stochastic models of income determination that reproduced the stylized fact that incomes follow a Pareto distribution.[16] Notably, this whole

[16] Champernowne's (1973) book was based on his undergraduate research completed in 1936 but updated and republished, and thus reflects this older approach to the question of the distribution of income.

research program eschewed analysis of the characteristics of particular individuals and focused instead on producing a model that matched the observed shape of the whole distribution.

In the same period that statistical research on the distribution of income took off (1890–1960), mainstream economic theory became increasingly dominated by an approach to understanding wages and productivity known as the "marginal productivity theory of distribution" (Pullen 2009). In general, "marginalism" refers to an understanding of economic decision-making in terms of the "margin"—how much is one more unit of a good worth? How much more could a business produce if it hired one more worker (or paid for one more hour of work)? The marginal revolution dates to the 1870s with nearly simultaneous publication of works in England (Jevons), France (Walras), and Austria (Menger) espousing some form of marginalist analysis (Mirowski 1984). Although these authors differed dramatically in their methodological approaches and substantive conclusions, their insights were eventually unified into a seemingly coherent school of economics called marginalism (Blaug 1972).

One of the most important substantive conclusions derived from marginalist analysis was the marginal productivity theory of distribution, developed by the American economist John Bates Clark. Writing at the end of the 19th century, Clark used calculus to conceptualize the marginal product of labor in terms of its partial derivative (Pullen 2009). In the context of the distribution of personal income, and assuming a competitive market economy, marginal productivity theory implies that an individual's wage will be a function solely of that individual's productivity. This theoretical link between wages and productivity persists into the present and forms the backbone of most neoclassical analysis of wages and labor markets. In macroeconomics, such approaches were linked to the Cobb-Douglas regressions. In microeconomics, the marginal productivity theory was transformed in the mid-20th century with the development of human capital theory. Rather than explaining the statistical distribution of all incomes in terms of some underlying law, human capital theories would seek to explain differential productivity between groups of workers based on their education and work experience.

In 1958, Jacob Mincer published an article based on his doctoral dissertation, titled "Investment in Human Capital." This article is widely regarded as launching the human capital turn in labor economics (McNulty 1980; Teixeira 2007). In it, Mincer explicitly frames his research against the older, statistical models of personal income distribution: "Economists have long theorized about the nature or causes of inequality in personal incomes. In contrast, the vigorous development of empirical research in the field of personal income distribution is of recent origin. Moreover, the emphasis of contemporary research has been almost completely shifted from the study of the causes of inequality to the study of the facts and of their consequences for various aspects of economic activity, particularly consumer behavior" (Mincer 1958, p. 281).

By "study of the facts," Mincer refers to the empirical estimation of the distribution of incomes associated with Pareto and Champernowne. Research on the "consequences for . . . consumer behavior" includes the short-lived line of research in macroeconomics connecting the personal distribution of income to the marginal propensity to consume. Mincer argued that economists needed to focus more attention on the causes of the distribution of income and to connect such research to the paradigm of rational choice and marginal productivity. In other words, Mincer argued that economics could understand the distribution of income as a consequence of the choices individuals make about how to invest in their own productivity, especially by seeking formal education or on-the-job training. In a 1997 retrospective, Mincer also connected the rise of human capital theory to the inability of 1950s macroeconomics to adequately account for the sources of economic growth (Mincer 1997; see also Teixeira 2007).

Soon, Mincer was joined by Theodore Schultz and Gary Becker in leading the charge for human capital theory. In 1962, Schultz and Becker produced a special issue of *Journal of Political Economy*, which brought human capital theory squarely into the limelight.[17] Blaug (1976, p. 827) documented the dramatic takeoff of human capital theory studies in the 1960s and 1970s, finding 800 articles on human capital theory published by 1966, and 2,000 articles by 1976. Blaug characterized the dominant form of research in the human capital approach in terms of its exploration of the "earnings function": "The [human capital research] program adds up to an almost total explanation of the determinants of earnings from employment, predicting declining investment in human-capital with increasing age, and hence lifetime earnings profiles that are concave from below. No wonder the bulk of empirical work inspired by the human-capital framework has taken the form of regressing the earnings of individuals on such variables as native ability, family background, place of residence, years of schooling, years of work experience, occupational status, and the like—the so-called 'earnings function'" (Blaug 1976, p. 832).

Several important observations follow for our story. First, human capital theory was interested in *earnings* or wages, not in all sources of income. Second, human capital research was interested in explaining variation in individual earnings in terms of a small set of important covariates, including

[17] There is a bit of a priority dispute over which of Mincer and Schultz should receive more credit for human capital theory. Some of Schultz's work in the 1950s makes similar conceptual moves to Mincer's dissertation (e.g., Schultz 1950). That said, commenters agree that the 1962 special issue marked the widespread takeoff of the approach.

especially education. Third, the human capital approach virtually invented the idea of "earnings regressions." As strange as it may seem in retrospect, before the 1960s, economists did not run—and were largely not capable of running—regression models predicting individuals' earnings with individual-level covariates. In part, economists avoided earnings regressions because the variations in individual earnings were not seen as an important theoretical topic (as opposed to, say, the aggregate labor share). But economists were also limited by the relative paucity of data. Fortunately, just as the human capital revolution was underway in economic theory, the microdata revolution was taking place.

The rise of the large-scale representative survey is a relatively recent phenomenon (Igo 2007). The first major representative surveys took place during the Great Depression, such as the famous Gallup Poll predicting Roosevelt's 1936 electoral victory over Landon. The technique expanded dramatically in the 1940s, including the 1940 launch of the "Monthly Report of Unemployment," which became the influential CPS in 1948 (Bureau of Labor Statistics 1984, p. 8). Although the CPS was born as part of the macroeconomic knowledge infrastructure (and remains the key tool for tracking unemployment), its structure and kernel (i.e., a large, nationally representative household sample) were relatively easy to extend to accommodate the needs of human capital researchers. The CPS, along with a handful of other large-scale, government-funded, representative surveys like the PSID (launched in 1968), would become the tool of choice for labor economists interested in understanding not just the overall shape of the income distribution, but also the covariates that predicted income (age, education, gender, and race). The availability of individual-level microdata contributed to the shift in the field of labor from studies of the demand for labor (firms) to studies of the supply of labor (individual laborers), and especially to the success of human capital theory and its earnings regressions. According to Stafford (1986, p. 388), in the 1970s, "about two-thirds of labor articles in major journals were on the broad subjects of labor supply and wage determination," and by the early 1980s, about two-thirds of empirical labor economics articles in top journals used individual-level microdata, and half of those drew on just three surveys (the CPS, the PSID, and the National Longitudinal Surveys; Stafford 1986, p. 395).

Surveys fit the research program of human capital theory nicely. Surveys were less effective than tax data at capturing financial income and profits but were relatively good at measuring earnings from labor. And large-scale demographic surveys like the CPS had the significant advantage of including all the covariates a researcher could want, especially education variables. As human capital theory increased in influence, its interests became even better reflected in the data collection practices of government surveys:

> Prior to the advent of human capital models of lifetime earnings, most sets of microdata did not have much information on work histories of individuals. Early work on earnings functions was commonly based on years of potential labor market experience, defined in terms of age and years of formal schooling. Because the theory emphasized the importance of on-the-job training through various types of job market experience, new and on-going data collection efforts obtained extensive information on job market experience. Variables such as years of full time experience, years of part time experience, and years in military service became widely available in cross-sectional data. (Stafford 1986, p. 398)

Conversely, because advances in survey-based microdata were driven by pressure from labor economists, and labor economists did not particularly care about non-wage income, improving measurement of such income was not a priority. Similarly, labor economists interested in the returns to a college degree or on-the-job training were satisfied with surveys that captured the middle 90% of the population accurately at the expense of known problems with income reporting at both the top and bottom of the distribution. Top incomes were especially poorly captured due to the problem of top coding. Thus, for researchers relying on the CPS, incomes at or above the 97% (approximately) were simply unobservable, and researchers either ignored these incomes or made static imputations about their size. For example, a review paper by the influential labor economists Katz and Autor (1999, p. 1471) assumed that there were was no trend in top incomes in order to impute their size (which also serves as a good example of how the consensus of the field at the time did not recognize the large increase in top incomes).

By the late 1970s, then, the modern knowledge infrastructure for labor economics had been assembled. Routinely conducted, large-scale, government-led or government-funded surveys produced individual-level microdata that provided good measurements of labor earnings and theoretically important covariates, but poor coverage of top incomes. These data were analyzed with the goal of understanding how individuals chose to invest in education and job training over the life course, and to measure the returns to a college degree, along with other predictors of individual earnings.[18] That surveys were technically incapable of measuring top incomes was simply not seen as a pressing problem.

From First Priority to Orphan Estimates

Contemporary discussions of the microdata revolution in empirical economics sharply contrast microdata with older, aggregate statistics, including national income accounts. Historically, the situation is a bit messier. During the 1950s

[18] These data were also well suited to other questions of policy concern including those asked by sociologists studying racial and gender gaps in earnings (Leicht 2008).

to 1970s, government statisticians worked hard to integrate microdata and the national accounts in order to produce comprehensive statistics on the distribution of income that would be more useful in combination. Economists in government in the 1940s considered combining these statistics to be their number one statistical priority. Yet, when the BEA began to publish such data in the 1950s, they were barely noticed. When funds to update and maintain the data could not be secured, the BEA ended the estimates, again, to little fanfare. What happened? Here, I trace one possible way that the personal distribution of income could have been incorporated into macroeconomists' knowledge infrastructure so that changes in top incomes would have been more visible. In doing so, I highlight the forces that produce inertia in the inequality knowledge infrastructure.

Following World War II, Congress's newly established Joint Committee on the Economic Report (later renamed the Joint Economic Committee, or JEC) laid out a series of "statistical gaps" that federal statistical agencies were charged with filling. First on this list was information on the personal distribution of income (JEC 1949, p. 86). In the 1950s, the BEA turned its resources to this task and published estimates for the distribution of income in the United States from 1944 to 1950 (BEA 1953; see fig. 7 for a graphical presentation of the 1950 estimate). By this time, national income data reports were published quarterly and received news coverage, much as they still do today. But in the 71 *New York Times* stories that reference the BEA in 1953 and 1954, just one mentioned the income distribution study. And it emphasized the gains in *average* income, discussing the distribution of income in just a few sentences at the end.[19] The BEA report itself argued for the importance of distribution data in "market analysis," with little mention of academic or policy implications.

Later efforts by the BEA to produce data on the distribution of income met with similarly little interest. The BEA produced annual estimates of income distribution from 1955 to 1964, and sought funds unsuccessfully from Congress to update the series and to better integrate survey data from the Census and Federal Reserve. Despite being deemed the most accurate and comprehensive available measurements of the distribution of income by a report to the JEC (T. Schultz 1964), and despite personal income distribution having been the primary statistical priority as recently as 1949, Congress denied a request for $60,000 per year to fund four permanent positions dedicated to the income distribution data. These positions were necessary to deal with the practical difficulties of bringing together multiple data sources, from dealing with errors on data tapes to tackling the problem of matching

[19] Charles E. Egan, "Family Income up $850 in Six Years," *New York Times*, October 12, 1953. The search term used here is *Office of Business Economics*, as the BEA was then known.

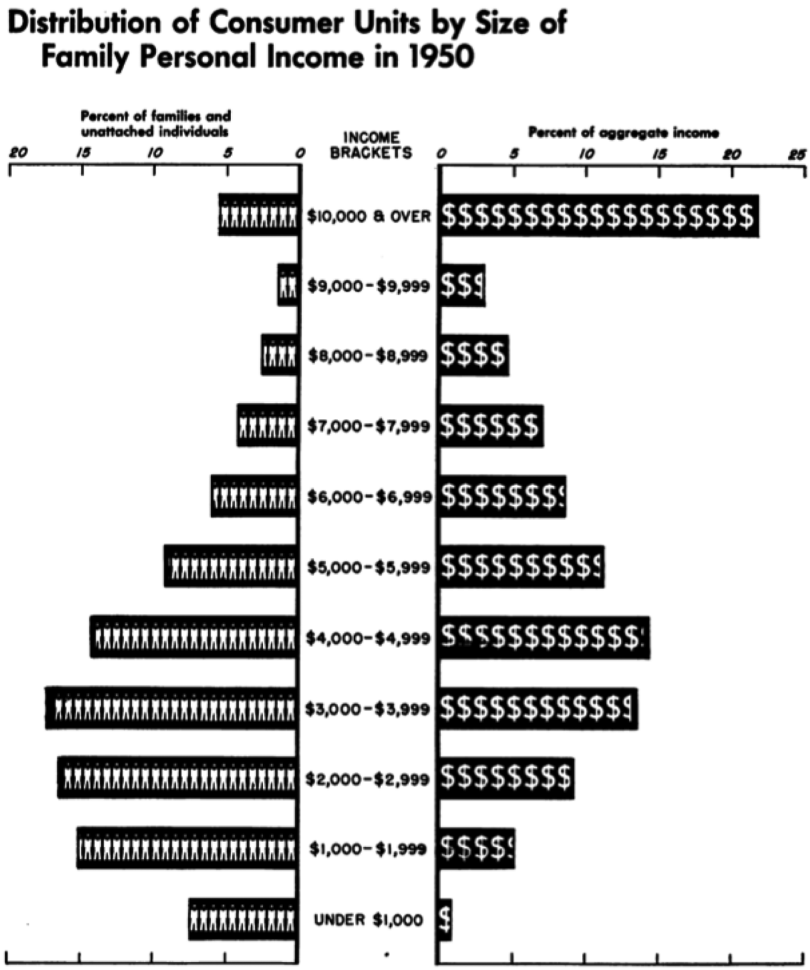

FIG. 7.—A graphical representation of the distribution of pretax income in the United States in 1950 (BEA 1953, p. 4).

households and tax units given the incredibly sparse information available in the tax record (see Budd 1971 for details). This request was, in part, badly timed: the Johnson administration faced strong competing demands throughout the mid-1960s that led to an effort to produce cuts and find novel sources of revenue even as the administration pushed expensive policy priorities including funding Great Society programs and the war in Vietnam (Quinn 2017). The BEA tried to link its research to the War on Poverty but failed to sway skeptical members of Congress. This failure followed the typical pattern of research on poverty being kept intentionally separate from discussions of inequality to avoid political blowback (O'Connor 2002).[20] The BEA would have been hard-pressed to point to massive interest from outside; for example, the *New York Times* makes no mention of the BEA's 1963 or 1964 publications estimating the distribution of income. And with the declining importance of institutional economists, no significant group of academics rallied to defend these data which were not seen as especially valuable by neoclassical macro and labor economists. The series then stopped abruptly at the end of 1964.

In 1974, the BEA resumed its work on income distribution. Unlike those by Piketty and Saez (2003), the BEA's estimates included some sociodemographic variables, as they combined data from the CPS alongside tax data and other sources (Radner and Hinrichs 1974). Thus, in some sense, these data could have been useful to labor economists interested in "gap research." It is difficult to speculate exactly why labor economists did not draw on these data, but we can see how their priorities were already better served by survey data—the BEA had a few demographic covariates, but no detailed work histories, and the BEA's major accomplishment was to reconcile survey-based measures of income with macroeconomists' particular definition of national income as embedded in the NIPA. But labor economists, interested primarily in *earnings* and not total income, would have little to gain from these reconciled measurements. Once again, the BEA's estimates received little to no public reaction (e.g., no mention in the *New York Times*). The 1974 estimates were not repeated in the following year, and the BEA did not publish further estimates of the distribution of income again until 2020, as discussed below.

By the end of the 1970s, the federal government had flirted with producing income distribution estimates compatible with the National Income and Product Accounts, but these efforts met with little interest from scholars and the public and were discontinued. The practical difficulties of merging income tax data and survey data, the relative disinterest in the academic community in such data, and the policy priorities of the federal government together prevented expansion of the existing knowledge infrastructure. What remained were survey-based measures derived from the U.S. Census's CPS and similar surveys conducted by the Bureau of Labor Statistics and the Federal Reserve to measure prices and consumption. These surveys captured the bulk of income—although with recognized problems of underreporting at the top and bottom[21]—and were ideally suited to the

[20] Congressional Hearing Transcripts, Budget Hearings for the Department of Commerce, 1965; see also Lebergott (1971, p. 117).

[21] The BEA argued before Congress that the Census surveys captured only about 90% of all income, due to problems of underreporting, top coding, and some categories that were simply not measured. Congressional Hearing Transcripts, Budget Hearings for the Department of Commerce, February 18, 1964.

research agendas of those interested in sociodemographic forms of inequality. These statistics did not, however, accurately measure all forms of income (such as capital gains), and they suffered from top coding, which made them largely incapable of tracking movements at the top of the income distribution. In terms of data, economic growth and income inequality were unlinked—no data set tracked both in comparable ways. This disconnect in the data used to study inequality and growth came just as a 30-year span of decreasing inequality would itself come to an end.

THE STYLIZED FACTS OF INEQUALITY

Having traced the construction of two disjoint knowledge infrastructures used to study incomes, I can now explain how economists knew what they knew about inequality in the 1980s–90s and how they were capable of missing new trends in top incomes.

In the 1980s and 1990s, labor economists both noticed and attempted to explain increases in inequality in the United States. The dominant characterization of this increase focused on the growing gap between the wages of college-educated and non-college-educated workers. One influential, and reasonably representative, example is Bound and Johnson (1992). They document the stagnation of average wages in the 1980s as well as dramatic shifts in the relative wages of individuals. Examining CPS data, they measure changes in the relative wages of individuals based on years of work experience, education, and sex. In the end, they conclude that "skill-biased technological change" drove major increases in the wage gap between college and non-college-educated workers. The emphasis here, as in much of the literature, is specifically on wage inequality, understood as an indicator of productivity (following from the marginal productivity theory of distribution and human capital theory).

Later work in the 1990s would debate the relative influence of international trade and globalization as alternatives to technological change as explanations for the observed gaps, as well as the persistence of race and gender gaps (see Card and DiNardo [2002] for one summary of the literature). While labor economists in the 1990s recognized the increasing right skew in the income distribution, they did not link it specifically to the growth of incomes at the very top, nor did they marshal data capable of explaining—or even *revealing*—movement at the very top.

Similarly, work in the same period by sociologists interested in inequality focuses on persistent wage gaps. Leicht (2008, p. 238) summarizes sociological research on inequality in the 1990s as "dominated by studies of racial and gender gaps" in contrast with economists' focus on education, technology, and trade. While sociologists' empirical focus and theoretical interests differed somewhat from those of labor economists, both disciplines relied on the same knowledge infrastructure, especially large-scale household surveys. An influential and representative example is Morris and Western (1999) that, like much work by economists, relies primarily on the CPS. For labor economists and sociologists interested in inequality, the growth of top incomes was almost invisible because the main knowledge infrastructure they relied on did not measure it.

The focus of inequality researchers on "gaps" was mirrored in popular coverage of inequality during the 1980s–90s. McCall (2013, pp. 53–95) analyzes news coverage of inequality in prominent newsmagazines during this period and found it to be very fragmented. Explicit discussions of economic inequality or the overall distribution of income are largely absent. Instead, like labor economists, public discourse focused on earnings disparities and education: "[The] earliest debates about income inequality seeped into the media as researchers began to notice the increase in the 1980s. Because earnings are the main component of income, and earnings disparities appeared to be rising in the 1980s, researchers focused at the outset more on the widening disparity in hourly *earnings* between high-skilled and low-skilled workers in the *labor market* than on disparities in *incomes* between rich and poor *families*" (McCall 2013, p. 59; original emphasis).

McCall did not find an increase in discussions of inequality in the popular press as inequality itself increased throughout the 1980s and continued through the early 2000s. Her findings are consistent with my analysis of newspaper articles (see fig. 3 above), which show that discussions of income inequality were roughly stable in the late 1980s to early 2000s. Rather than an increase in the *amount* of inequality discourse, she documents a change in the nature of inequality discussions in the early 2000s, which shift from a focus on earnings and the labor market to a focus on incomes and tax policy. "Remarkably, the role of tax policies in affecting top incomes did not emerge as a focal part of the academic story until the early 2000s, when new 'facts' about income inequality became available" (McCall 2013, p. 59). Popular discussion here followed the trends in academic research. In the 1980s–90s, labor economists identified earnings disparities as the novel stylized fact of interest, and public discourse focused on issues like the minimum wage, trade, and education. In the 2000s, the identification of the stylized fact of increasing top income inequality would alter the landscape of public debate about inequality.

Macroeconomists, for their part, mostly ignored the personal distribution of income in favor of continued debate about the stability of labor's share of national income and discussions of the big three macroeconomic variables: growth, inflation, and unemployment. Dominant macroeconomics textbooks, for example, included lengthy discussion of labor's and capital's respective shares of income, but no discussion of income inequality. Mankiw's (2003, p. 173) intermediate textbook is an influential exemplar: the entire book contains just one reference to economic inequality, and it is in the context of unemployment. Labor's share, on the other hand, is discussed and even modeled prominently in the contexts of the Cobb-Douglas production function (pp. 72–73) and growth accounting models (pp. 229–31). Labor's share was also a continued object of research, from debates about how best to measure it (Krueger 1999) to empirical observations of its decline starting in the 1980s (Gordon 1988).

Some heterodox economists devoted more space to income inequality in their textbooks, but they too relied on the knowledge infrastructures developed by mainstream economists and thus inherited those priorities and omissions. For example, heterodox economist Edward Wolff (1997) published an entire textbook entitled *Economics of Poverty, Inequality and Discrimination*. Strikingly, Wolff's chapter on historical trends in inequality begins by reporting top income shares from the 1910s to 1950 (citing data from the Commerce Department), but then switches to over to Gini coefficient measurements from the CPS starting with the 1950s without mentioning the dearth of available information on top incomes in the 1960s–90s (Wolff 1997, pp. 70–74). The 2008 revision of the textbook, *Poverty and Income Distribution*, added an extensive discussion of Piketty and Saez's (2003) work on top income shares.

In sum, the stylized facts of inequality in the 1980s and 1990s included an impressive and important list of findings: the widening gulf between college-educated and less educated workers, the persistence of race and gender gaps, and the overall decline in labor's share of national income. But, in general, the two largest groups of economists potentially interested in income inequality relied on knowledge infrastructures that were built for other theoretical and empirical purposes; thus they were incapable of seeing changes in top incomes. Other scholars, including sociologists and heterodox economists who made use of those same infrastructures for different theoretical purposes, were subject to the same constraints.

Krugman and the 1992 Presidential Election

Although most sociologists and economists conceptualized and analyzed inequality in ways that left the growth of top incomes of the 1% invisible, there were notable exceptions. Economists specifically interested in taxation paid attention to the IRS's Statistics of Income. Drawing on these data, Feenberg and Poterba (1993) noted the growth of incomes of the top 0.25%. Their paper appears to have attracted some notice from fellow tax scholars in the 1990s—and later from Piketty and Saez, who cite the paper as a precursor—but it had little resonance in the broader economics literature.[22] Feenberg and Poterba (2000) would repeat their argument that inequality scholars failed to pay sufficient attention to the very top incomes through the 1990s, and they explicitly noted the problem of top-coded survey data as an impediment to studies of this income group.

More surprising than this academic outlier is Paul Krugman's highly public 1992 calculation of the income gains of the top 1%. In the early 1990s, Paul Krugman was a noted international trade theorist in the process of transitioning to his current status as a leading public intellectual. His first major book for popular audiences, *The Age of Diminished Expectations*, was published in 1990 and included a very short chapter on income distribution, noting the growing wage gap between more and less educated workers and offering a brief analysis of the increased incomes of the top 10% (Krugman 1990, pp. 19–25). In an interview, Krugman (1998) later noted that his editors objected to this chapter's inclusion: "For complicated reasons, that book was initially a *Washington Post* project. And my editors at the *Post* tried to pressure me into taking the income distribution chapter out, saying that nobody cared about that issue."

Two years later, Krugman published a more surprising finding. Based on a Congressional Budget Office (CBO) publication with information derived from tax data, Krugman argued that between 1977 and 1989, 60% of the income gained by the country went to the top 1% of income earners. This finding was popularized in the *New York Times*, criticized by the *Wall Street Journal* and the Bush Treasury Department, and featured in stump speeches by then presidential candidate Bill Clinton. The CBO eventually responded to the critiques with a study of its own that supported Krugman's overall finding, but with a caveat that the number dropped a bit when adjusted for the size of families (a common practice at the CBO, though not common in many other income data sets). One *New York Times* summary of the debate is particularly revealing:

> One reason that the issue remained relatively invisible for so long—in spite of broad agreement among academic economists that income and wealth have grown more unequal—is that statistics on distribution are among the mushiest and murkiest in economics. *There are literally dozens of ways to slice and dice the data, not to mention hundreds of different data series*. Confronted by Republican legislators after Governor Clinton started peppering his speeches with

[22] According to Google Scholar (accessed 3/29/2021), Feenberg and Poterba (1993) was cited just 66 times through 2000 (seven years after publication), primarily by articles on the effects of tax reforms. It was then cited another 312 times between 2001 and 2020 after it was cited by Piketty and Saez (2001)–the working paper version of Piketty and Saez (2003). In contrast, Piketty and Saez (2003) was cited 66 times already in the year it was published and 995 times by 2010 (seven years after publication). I thank a reviewer for suggesting this comparison.

> the statistic on the top 1 percent, the budget office was called upon to assess the Krugman calculation. Several weeks later, it issued a report that gave a number of alternate measures of the gains by the top 1 percent. Every measure showed that the top 1 percent of families reaped an outsized share of the gains. By one calculation 70 percent of the rise in average after-tax family income went to the top 1 percent, rather than the 60 percent figure that Mr. Krugman had initially estimated and that Governor Clinton has been using.[23] (emphasis added)

Krugman's 1992 calculation represents a significant challenge to the initial premise of this article—that the growth of the top 1% was a surprising finding in the early 2000s. Clearly, there was some political awareness in 1992 that the incomes of the top 1% had grown dramatically. Yet, despite the political salience of Krugman's calculation during the 1992 election, the stylized fact that the top 1% had captured most of the income gains of the 1980s and grown to historically unprecedented levels did not capture sustained attention. Without a recognized place among academics, or a tight connection to a particular knowledge infrastructure, it became just one more talking point bandied about in a presidential election. One piece of evidence for this claim (in addition to the discussion above about the consensus stylized facts about inequality in the period) is Krugman's own later academic work. In 1995, Krugman published a foray into the debate about the role of trade in increasing inequality. Krugman (1995) makes no mention of the growth of the top 1%, instead sticking to the mainstream academic debate about the increasing education wage gap and concluding that increased world trade was not then responsible for growing income inequality. Finally, Krugman's calculation also focused on a very narrow time window—1977–89—corresponding to the period covered by the CBO. It would be Piketty and Saez who would go back to the IRS data and construct a new knowledge infrastructure that put contemporary top incomes into a much longer historical and comparative perspective.

REDISCOVERING THE 1%

In 2001, Piketty and Saez released an NBER working paper containing estimates of the share of income accruing to the top 5% and 1% of income earners in the United States from 1913 to 1998. Their paper relied heavily on tax data, combined with NIPA data to determine some of the relevant denominators (e.g., total wage income). Their estimates differed from Feenberg and Poterba (1993, 2000) in five interesting respects. First, Piketty and Saez estimated a consistent series back to 1913, several decades earlier than Feenberg and Poterba. Second, in order to extend their series back further, Piketty and Saez used a slightly more complicated method to generate their denominators

[23] Sylvia Nassar, "Richest Getting Richer Now Top Political Issue," *New York Times*, May 11, 1992.

so that they could take into account the much smaller percentage of income that was reported on tax returns in earlier periods. Third, Piketty and Saez built on Piketty's recently completed work on top incomes in France, where top income shares had stayed flat in the 1980s and 1990s. This comparison suggested that the U.S. story was explainable by factors specific to American political economy and not a universal feature of late-20th-century economies. Fourth, Piketty and Saez noted that the top 1% increasingly derived their incomes from wages and salary, and not simply from capital gains, thus rejecting a sharp division between their work on overall income inequality and the work of labor economists focused solely on wage income. Fifth, Piketty and Saez made their data freely and easily available. In the 2000s, this sharing took the form of hosting and updating a simple spreadsheet on Saez's faculty homepage.[24] This spreadsheet made it easy for other researchers not only to check Piketty and Saez's work, but also to build on it and incorporate it into other analyses. Even before Piketty and Saez and other collaborators established a more formal knowledge infrastructure (as discussed below), their work was already much more accessible than the data housed on tapes in the federal statistical agencies in the 1960s and 1970s (Budd 1971). All told, Piketty and Saez produced and shared a data set capable of making the provocative claim that the share of income accruing to the top 1% in the United States had returned to levels not seen since the Great Depression.

Although these five features help to explain the portability and appeal of Piketty and Saez's work as compared to earlier discussions, it remains hard to say definitively why it had such a large and sustained impact (as depicted back in fig. 5). That their data series extended back to the Great Depression may have played a large role in capturing attention. Additionally, the early 2000s saw researchers interested in more conventional wage inequality casting about for new explanations. Katz and Autor's (1999) review of the literature on changes in inequality in the United States noted some evidence that the top incomes had grown dramatically (even as their own quantitative evidence, relying on CPS data, analyzed only incomes up to the 97th percentile). Specifically, Katz and Autor (1999, p. 1468) cited growing evidence on the tremendous growth of CEO and athlete pay (e.g., Hall and Liebman 1998). It is possible that Piketty and Saez's work caught on academically because it helped make sense of increasing evidence from the labor economics literature that something interesting was happening at the top of the income distribution that had not previously been fully measured and did not easily fit within the framework of human capital theory.

[24] The most recent update to this Excel file as of the time of this writing was posted in February 2020 and contains data through 2018. See https://eml.berkeley.edu/~saez/TabFig2018.xls (accessed 3/30/2021). I thank a reviewer for suggesting this discussion.

Piketty and Saez also benefited from early and positive attention from the media.[25] Specifically, Paul Krugman, by then a columnist at the *New York Times*, wrote an extensive magazine piece drawing heavily on their research. Krugman (2002) argued that the United States had entered a "new Gilded Age." Krugman summarized 15 years of CPS-based studies that showed increases in income in the top 5%, starting in the late 1970s.[26] But Krugman then went on to claim that studies that relied on categories as big as the top 10% or 5% missed the big change in the new Gilded Age:

> Most of the gains in the share of the top 10 percent of taxpayers over the past 30 years were actually gains to the top 1 percent, rather than the next 9 percent. In 1998 the top 1 percent started at $230,000. In turn, 60 percent of the gains of that top 1 percent went to the top 0.1 percent, those with incomes of more than $790,000. And almost half of those gains went to a mere 13,000 taxpayers, the top 0.01 percent, who had an income of at least $3.6 million and an average income of $17 million. (Krugman 2002)

Krugman noted that the three main arguments (globalization, skill-biased technological change, and the so-called "superstar" hypothesis) economists had generated to try to understand growing income inequality in the 1980s–90s seemed inadequate in the face of the Piketty and Saez data: "Globalization can explain part of the relative decline in blue-collar wages, but it can't explain the 2,500 percent rise in C.E.O. incomes. Technology may explain why the salary premium associated with a college education has risen, but it's hard to match up with the huge increase in inequality among the college-educated, with little progress for many but gigantic gains at the top. The superstar theory works for Jay Leno, but not for the thousands of people who have become awesomely rich without going on TV" (Krugman 2002).

Instead, Krugman argued, economists would need to bring in social and political factors, like social norms, to make sense in the massive increases in executive pay that helped to drive the incomes of the 1%. Other economists, sociologists, and political scientists would build on Piketty and Saez's findings to make claims for the importance of politics (Hacker and Pierson 2010), financialization (Kaplan and Rauh 2007), the rise of agency theory, the death of unions, and more. In the wake of the 2008 financial crisis, as social movement

[25] Critic Alan Reynolds (2007:2) claimed: "After 35 years of writing on economic issues, I do not recall any other private and unofficial estimates that were as widely and uncritically repeated as the Piketty-Saez estimates on income shares of the top 1 percent . . . Searching Google for 'Emmanuel Saez' in early October turned up 51,700 entries, including 871 that also involved the *TheNew York Times*. Similar searches yielded 814 joint references to Saez and the *Washington Post*, 568 for *TheWall Street Journal*, 375 for the *Financial Times*, and 319 for *USA Today*."

[26] Krugman makes no mention of his own earlier 1992 calculations on the growth of the top 1%, showing just how tangential they had been to the larger conversation on inequality in this period.

activists and progressive politicians sought to make sense of growing inequality, they were able to tap into these new theoretical debates and the data underlying them. Senator Bernie Sanders rose to national fame when he gave a widely publicized 8.5 hour speech (later published as a standalone book) as part of a filibuster against a proposed tax bill in 2010. In that speech, Sanders used the phrase the "top 1 percent" 26 times, citing data on the increasing concentration of income and wealth (Sanders 2012). Most famously, in 2011, Occupy Wall Street explicitly mobilized in the statistical language of top income shares with the slogan "We are the 99%" (Gould-Wartofsky 2014). Sanders's speech and Occupy Wall Street's slogan led to a spike in public discussions of the 1% in 2010–11 (fig. 4 above). Piketty (2014) himself would write a best-selling book connecting the growth of top incomes to newly discovered "fundamental laws of capitalism" relating the returns on capital, the rate of economic growth, and concentration of income at the very top; a sequel focused on the role of ideology in justifying inequality across time and societies (Piketty 2020). The stylized fact of the growth of the top 1% had entered academic and political debates in force and lent support to new and different theories of inequality.

Not all economists accepted Piketty and Saez's findings uncritically. Conservative think tanks continued their war on all findings of increased income inequality, much as they had contested Krugman's 1992 calculation. For example, a critique published in the Cato Institute's journal *Policy Analysis* claimed that Piketty and Saez dramatically overstated the rise of the top 1% by misusing tax data and that their estimates were inconsistent with the story told by survey data: "If the Piketty and Saez estimates actually demonstrated a continuous and credible upward trend toward greater inequality since the late-1980s, all other estimates of income distribution would have to be wrong—including those of the Census Bureau, the CBO, and the Federal Reserve Board" (Reynolds 2007, p, 18).

Piketty and Saez showed a change in recorded, taxable income accruing to the top 1%, but Reynolds argued that this observation was consistent with a shift in *how* the 1% earned their money rather than the total *amount* the top 1% earned: top earners responded to changes in the tax law (lower top marginal tax rates) by shifting income from corporate tax filings to individual tax filings: "Studies based on tax return data provide highly misleading comparisons of changes to the U.S. income distribution because of dramatic changes in tax rules and tax reporting in recent decades. Aside from stock option windfalls during the late-1990s stock-market boom, there is little evidence of a significant or sustained increase in the inequality of U.S. incomes, wages, consumption, or wealth over the past 20 years" (Reynolds 2007, p. 1).

Piketty and Saez (2006) responded to Reynolds point-by-point, noting that the discrepancies between their findings and survey-based measures are completely logical, and in fact, the main novelty of their study: "The reason for the discrepancy is that the Census Bureau estimates are based on survey data which are not suitable to study high incomes because of small sample size and top coding of very high incomes. In contrast, tax return data provide a very accurate picture of reported incomes at the top. Our key contribution was precisely to use those tax data to construct better inequality estimates." My point here is not to claim that Piketty and Saez are right and Reynolds is wrong, or vice versa. Rather, I aim to document just how much was still up in the air in the mid-2000s in terms of the measurement of inequality (and indeed, even now, see, e.g., Auten and Splinter [2019], Kopczuk and Zwick [2020], Saez and Zucman [2020], and Guyton et al. [2021]). Although extensive debates still exist about the measurement of unemployment, inflation, and economic growth, the existence of widely accepted, bureaucratically produced, official, and timely measures institutionalized in devices forestalls many of the sorts of challenges present in both the 1990s "Krugman calculation" debate and in the critical reception of Piketty and Saez. Basic definitional issues—what kinds of income should be studied (wages, capital gains, nonmonetary remuneration, before tax or after tax and transfer, etc.), the units of analysis (families, households, individuals, "tax units"), and the relevant metrics for comparison (top 1% or 0.1% income shares, 90/50 or 90/10 ratios, Gini coefficients)—are all still live controversies in the study of income inequality.

Inequality is not unknown, nor unknowable. Piketty, Saez, and more than 100 collaborators continue working to construct new knowledge infrastructures in the form of the World Inequality Database, which aims to produce standardized measures of top incomes for different countries (Alvaredo et al. 2018) as well as distributional national accounts that would harmonize data on the personal distribution of income with the data used to measure each country's GDP (Piketty et al. 2018). These efforts are explicitly aimed at pressuring governments to take over the production of official inequality statistics (Saez and Zucman 2020, p. 54). At the same time, the Washington Center for Equitable Growth (WCEG) has been working to bring inequality back to the theoretical center of economics by funding and promoting research, including revisiting the role of the distribution of income in shaping the marginal propensity to consume (Fisher et al. 2018), a central question in macroeconomics. The WCEG has also been instrumental in pushing the BEA to produce new measures of income distribution (Boushey 2018).[27] In early 2020, the BEA published provisional data on the distribution of personal income based on surveys, administrative records, and tax records. These data include breakdowns for the share of income going to the top

[27] David Harrison, "Government to Measure Who Gets What Share of the Economic Pie," *Wall Street Journal*, April 16, 2019.

quintile, top 5%, and top 1%, although these statistics currently only cover 2007–16.[28] These efforts could transform calculations of top income from a contentious topic for debate into a routine production of an established knowledge infrastructure.

DISCUSSION AND CONCLUSION

How do past research priorities shape future knowledge production? Knowledge infrastructures are one important site where the tyranny of inertia (Bowker and Star 1999) manifests, channeling and limiting the attention of researchers. Theoretical, empirical, and political priorities are built into knowledge infrastructures; these continue to shape the kinds of data that are collected, processed, and distributed, which in turn makes some research more doable (Fujimura 1987) while rendering other topics out of focus (Peterson 2017). Knowledge infrastructures produce knowledge, making it possible to monitor changes and identify trends. But knowledge infrastructures can also produce ignorance.

As the case of the rediscovery of the 1% shows, making observations outside a knowledge infrastructure is difficult and dismissing those observations is more easily accomplished. The NIPA and surveys like the CPS make even small movements in GDP growth, unemployment, the gender wage gap, or the returns to a college degree impossible to miss. Monthly or quarterly reports track the main aggregates, and researchers constantly pore over the details, analyzing the causes of any changes. When income inequality began to grow in the 1980s, researchers spotted it very quickly, and public discourse reflected this rapid awareness. In contrast, top incomes fell outside of the dominant knowledge infrastructures of the 20th century, and thus researchers were capable of missing their growth. Public discourse mirrored this omission. In order to establish trends in top incomes, scholars in the 2000s and 2010s constructed a secondary knowledge infrastructure, building on existing data sources while simultaneously drawing attention to the limitations of those sources and pressuring the state for resources to expand efforts to track inequality and prioritize the routine collection of data on top incomes.

Extant knowledge infrastructures do not completely determine the direction of research—as Piketty and Saez's work shows. Yet this case offers several demonstrations of how existing knowledge infrastructures can lower or raise the costs of pursuing a given line of research. Here, I highlight two: debates in macroeconomics over the link between inequality and the fiscal spending multiplier and the influence of mainstream economics on heterodox

[28] https://www.bea.gov/data/special-topics/distribution-of-personal-income (accessed 11/8/2020).

economics. After some interest in the 1930s–40s, macroeconomists abandoned the question, determining that differences in spending behavior across the income distribution were not important enough to be included in their models, and in turn, that it was not worth the trouble to collect data on the distribution of income that was compatible with other macroeconomic aggregates like GDP. These decisions were locked into macroeconomic knowledge infrastructures, making it difficult for scholars to revisit these assumptions, even as the underlying distribution of income changed and as macroeconomic models grew increasingly complex. Macroeconomists only returned to the question after Piketty and Saez's work had drawn attention back to the distribution of income, and after the wave of Keynesian stimulus spending that followed the 2008–9 crisis (see Auclert and Rognlie 2020). Not until 2020 did the BEA begin publishing data that integrated the distribution of personal income with other macroeconomic aggregates.

Perhaps more surprising, the case also shows how dominant knowledge infrastructures can shape research trajectories of subordinate fields. The NIPA and large surveys like the PSID and CPS were designed largely by and for mainstream economists. Yet their decisions influenced research trajectories in heterodox economics and in sociology. These fields relied on the same infrastructures and thus were constrained by their limitations. For example, the 1997 edition of heterodox economist Wolff's textbook on income distribution included measures of top incomes for the first half of the 20th century, but omitted them starting in the 1950s, relying instead on calculations based on the CPS. The textbook was updated in light of the knowledge infrastructure that Piketty and Saez (2003) had begun to construct. Even those scholars whose theoretical inclinations led them to focus very explicitly on the distribution of income and income inequality outside a human capital or macroeconomic framework still relied on well-resourced, legitimated, mainstream knowledge infrastructures and thus incorporated some of those infrastructures' assumptions and limitations.

Beyond reinforcing existing theoretical claims that knowledge infrastructures are sites where the past shapes the present, I identify two further insights from this case in particular: the possibility for knowledge infrastructures to decay and the importance of sustained attention rather than isolated observations. The methods that economists used to track top income shares via tax data in the 1920s–50s were systematically abandoned. These methods were then rediscovered in the 2000s. This pattern suggests that, much like science itself, knowledge infrastructures do not necessarily follow a trajectory of monotonic improvement. Rather knowledge infrastructures can be abandoned or forgotten, making it possible for trends that would have been visible to become hard to see. Knowledge infrastructures can also suffer from a lack of maintenance or become less useful due to changes in the trends they are trying to monitor. In his newer work, Piketty (2020, p. 672) notes that wealth data across Europe and the United States have grown less useful because of the increased use of tax havens to hide wealth and the declines in the estate tax that produced much of the data.

This potential for decay is particularly important because knowledge infrastructures shape sustained attention, and this case suggests that patterns of sustained attention may be as important as flashes of insight for the trajectory of a field. Despite its absence in dominant infrastructures, the rise of top incomes was not completely impossible to spot—Krugman's 1992 calculation and work by Feenberg and Poterba (1993) offer two examples of economists who did notice the change. But even these calculations came a decade after increases in top incomes had begun, and they had little lasting impact. They did not fit into a broader theoretical framework for making sense of labor markets or macroeconomies, and they were not the outcome of knowledge infrastructures that provided sustained attention to the topic. Those aspects of the world that are left out of existing knowledge infrastructures—that do not receive sustained attention—are more capable of being ignored and require substantial infrastructural work to bring to collective attention. As an isolated finding, Krugman's calculation was a political claim useful for criticizing the economic legacy of President Ronald Reagan. In contrast, Piketty and Saez (2001, 2003) framed their calculation in the long sweep of the past century. They published their findings in academic journals and continued producing and distributing data on the top incomes, at first on their own websites and then as part of a large comparative and collaborative research program (Alvaredo et al. 2018). In short, Piketty and Saez built a new knowledge infrastructure for top incomes that both analyzed data and interpreted it (at least enough to produce new stylized facts), and thus provided sustained attention that made it difficult for top incomes to simply recede into the background as they did after the 1992 presidential election. When public interest turned toward inequality again in the wake of the 2008 financial crisis, Piketty and Saez's data, and a new body of inequality scholarship built upon them, were available to provide support for new political movements and policy proposals.

The case studied here should be understood as the beginning, not the end, of a research program on the dynamics of knowledge infrastructures. Future research on this and other cases would help tease apart some of the interwoven explanatory threads in this case including which factors—intradisciplinary and academic struggles, larger political dynamics, and material and practical limitations—were most responsible for the durability of existing knowledge infrastructures and the barriers to constructing new ones, as well as how the changes in technology (such as the reduced cost of producing and sharing data sets) have affected the stability of knowledge infrastructures over time.

Moving beyond the case of the 1%, the approach to inequality knowledge infrastructures presented here may also offer insights into other patterns in the history of inequality research. Here, I focus on persistently ignored calls for research integrating organizational perspectives. In their recent book, Tomaskovic-Devey and Avent-Holt (2019) reiterate Baron and Bielby's (1980) 40-year-old call for inequality research to integrate organizational perspectives, as organizations are the main sites where resources are accumulated and distributed. They argue that Baron and Bielby's program stalled for data reasons: "The critical new structuralist literature on dual and segmented labor markets foundered on the scarcity of quantitative workplace data . . . nearly the only data anyone had at the time were national, and sometimes local, surveys of individuals" (Tomaskovic-Devey and Avent-Holt 2019, p. 23). This absence of employer-employee matched data has severely constrained research efforts. For example, the stylized fact of the gender wage gap—the "80 cents on the dollar" figure that circulates widely—is derived from aggregate statistics from household surveys and does not measure same-job, different-pay discrimination, though it is often misinterpreted as doing so, creating confusion in public and political debates (Hirschman 2021). Smith-Doerr et al. (2019) note that only five quantitative papers on the gender wage gap in the United States (including their own) can actually compare men and women working in the same jobs at the same employer because of a lack of data, and none of these papers have nationally representative samples. Instead, researchers focus on special cases where data are available (like Smith-Doerr et al.'s study of federal science agencies). My account suggests an approach for understanding how it came to be that surveys dominated our inequality knowledge infrastructure and how this dominance may have created an "ignorance loop" (Durant 2020), where the absence of employer-employee matched data made it difficult for certain kinds of research to be conducted, shaping theories of inequality that deemphasized organizations and in turn reinforced the existing patterns of data collection that failed to collect organizational data relevant for inequality research.

This approach to understanding the dynamics of knowledge infrastructures and ignorance loops should also be useful for understanding the trajectories of scientific research beyond questions of income inequality. For example, sociologists have long noted a dearth of research in sociology on Native Americans across subfields (Snipp 1992; Huyser et al. 2010; Walter and Anderson 2013; Brayboy and Chin 2020). Although there are many reasons for this historical exclusion—including the centrality of settler colonial ideology of ignoring the continued existence of indigenous peoples (McKay, Vinveta, and Norgaard 2020)—this pattern is exacerbated by sociology's heavy reliance on household surveys that typically have very small samples of Native respondents. Researchers can overcome these problems, for example by gaining access to large-N administrative data (e.g., Akee, Jones, and Porter 2019). But doing so represents an additional barrier, and much research has instead simply left Native Americans out or collapsed Native respondents into an undifferentiated "other" racial category. This process in turn reinforces the ignorance loop: gaps in the infrastructure shape what research is done, the research that is done shapes priorities for maintaining and creating the infrastructure.

So far, I have drawn on examples that highlight the consequences of the design and implementation of survey-based knowledge infrastructure and, given the centrality of surveys to contemporary social science, these examples are not exhaustive. One could offer a similar reading of Pettit's (2012) analysis of how survey-based knowledge infrastructures overstated the economic progress of Black men in the 1990s because most surveys do not sample incarcerated people, and racialized mass incarceration sufficiently altered the denominators of many surveys as to induce the appearance of significant progress. But these dynamics are not limited to knowledge infrastructures rooted in surveys; future research would benefit from comparing examples from diverse infrastructures including censuses (e.g., Smith and Gates [2001] on the undercounting of same-sex couples in the 2000 census) and natural science contexts like climatology (e.g., Edwards [2010] on the relative absence of weather stations in Africa). Each of these empirical cases would require their own detailed analysis, but the framework laid out here— focusing on how knowledge infrastructures collect, process, and distribute data and how that enables the production of knowledge (and especially of trends and stylized facts), but also produces a tyranny of inertia where consequential initial decisions flowing from particular political and theoretical priorities shape the production of ignorance—should be helpful for teasing out the process.

Throughout this analysis I have focused on how knowledge infrastructures shape the trajectory of academic research, producing both knowledge and ignorance. Knowledge infrastructures may be just as important, though, for how they shape politics. Just as political priorities shape what kinds of data are collected, and what kinds of knowledge infrastructures are built and maintained, knowledge infrastructures may in turn shape politics in a democracy. Piketty and Saez's work resonated throughout the social sciences, but also in the form of mass mobilizations like Occupy Wall Street and elite political rhetoric. Piketty's (2014, 2020) blockbuster books have combined calls for new kinds of taxes with calls for new kinds of data—that is, one upside of a wealth tax would be the creation of new, high-quality wealth data. Even if the wealth tax ended up being too small to dramatically alter inequality, Piketty argues that the knowledge infrastructure that such a tax would help to create would provide fuel for antiplutocratic politics and policies, much as the rediscovery of the 1% altered the course of debates around income inequality.

REFERENCES

Akee, Randall, Maggie R. Jones, and Sonya R. Porter. 2019. “Race Matters: Income Shares, Income Inequality, and Income Mobility for All Races.” *Demography* 56 (3): 999–1021.

Alchon, Guy. 1985. *The Invisible Hand of Planning*. Princeton, N.J.: Princeton University Press.

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emannuel Saez, and Gabriel Zucman. 2018. *World Inequality Report*. Cambridge, Mass.: Harvard University Press.

Auclert, Adrien, and Matthew Rognlie. 2020. “Inequality and Aggregate Demand.” Working paper. Stanford University, Department of Economics. https://web.stanford.edu/~aauclert/inequad.pdf.

Auten, Gerald, and David Splinter. 2019. “Top 1 Percent Income Shares: Comparing Estimates Using Tax Data.” *AEA Papers and Proceedings* 109:307–11.

Barber, William J. 1985. *From New Era to New Deal: Herbert Hoover, the Economists, and American Economic Policy, 1921–1933*. Cambridge: Cambridge University Press.

Baron, James, and William Bielby. 1980. “Bringing the Firms Back In and Stratification, Segmentation, and the Organization of Work.” *American Sociological Review* 45:737–65.

BEA (Office of Business Economics). 1953. *Size Distributions of Income in the U.S., 1944–1950*. Washington, D.C.: Government Printing Office.

Biddle, Jeff. 2012. “The Cobb–Douglas Regression and the Measurement of Economic Growth and Its Causes.” Working paper. Michigan State University, Department of Economics.

Blaug, Mark. 1972. “Was There a Marginal Revolution?” *History of Political Economy* 4 (2): 269–80

———. 1976. “The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey.” *Journal of Economic Literature* 14 (3): 827–55.

Bokulich, Alisa. 2020. “Towards a Taxonomy of the Model-Ladenness of Data.” *Philosophy of Science* 87 (5): 793–806.

Bonilla-Silva, Eduardo. 2017. *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*,. 5th ed. London: Rowman & Littlefield.

Borgman, Christine L., Peter T. Darch, Ashley E. Sands, Jillian C. Wallis, and Sharon Traweek. 2014. “The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management.” Pp. 257–66 in *Proceedings of the 14th ACM/IEEE-CSJoint Conference on Digital Libraries*. September.

Bound, John, and George Johnson. 1992. “Changes in the Structure of Wages in the 1980’s: An Evaluation of Alternative Explanations.” *American Economic Review* 82 (3): 371–92.

Boushey, Heather. 2018. “Testimony by Heather Boushey before the Joint Economic Committee.” Hearing on Measuring Economic Inequality in the United States. October 16. https://equitablegrowth.org/testimony-by-heather-boushey-before-the-joint-economic-committee/.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass.: MIT Press.

Boyce, Angie M. 2016. “Outbreaks and the Management of ‘Second-Order Friction’: Repurposing Materials and Data from the Health Care and Food Systems for Public Health Surveillance.” Special Issue: Knowledge Infrastructures. *Science and Technology Studies* 29 (1): 52–69.

Brayboy, Bryan Mckinley Jones, and Jeremiah Chin. 2020. “On the Development of Terrortory.” *Contexts* 19 (3): 22–27.

Budd, Edward C. 1971. “The Creation of a Microdata File for Estimating the Size Distribution of Income.” *Review of Income and Wealth* 17 (4): 317–33.

Budnick, Jamie. 2020. “The New Gay Science: Sexuality Knowledge, Demography, and the Politics of Population Measurement.” Doctoral dissertation. University of Michigan, Department of Sociology.

Bureau of Labor Statistics. 1984. "The Current Population Survey: A Historical Perspective and BLS's Role." *Monthly Labor Review* 107 (June): 8–14.
Campbell, Andrea Louise. 2012. "Policy Makes Mass Politics." *Annual Review of Political Science* 15 (1): 333–51.
Card, David. 2011. "Origins of the Unemployment Rate: The Lasting Legacy of Measurement without Theory." *American Economic Review* 101 (3):552–57.
Card, David, and John E. DiNardo. 2002. "Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." NBER Working Paper 8769. National Bureau of Economic Research. Cambridge, Mass.
Carson, Carol. 1971. "The History of the United States Income and Product Accounts: The Development of an Analytical Tool." Doctoral thesis. George Washington University, Department of Economics.
Champernowne, David Gawen. 1952. "The Graduation of Income Distributions." *Econometrica* 20 (4): 591–615.
———. 1973. *The Distribution of Income between Persons*. Cambridge: Cambridge University Press.
Clemens, Austin, and Michael Garvey. 2020. "Structural Racism and the Coronavirus Recession Highlight Why More and Better U.S. Data Need to Be Widely Disaggregated by Race And Ethnicity." *Washington Center for Equitable Growth*. Issue Brief: Inequality and Mobility. https://equitablegrowth.org/wp-content/uploads/2020/09/092520-oversampling-racial-data-ib.pdf .
Collins, Robert M. 2000. *More: The Politics of Economic Growth in Postwar America*. New York: Oxford University Press.
De Vroey, Michel. 2016. *A History of Macroeconomics from Keynes to Lucas and Beyond*. Cambridge: Cambridge University Press.
DeRock, Daniel. 2021. "Hidden in Plain Sight: Unpaid Household Services and the Politics of GDP Measurement." *New Political Economy* 26 (1): 20–35.
Downey, Greg, Kristin R. Eschenfelder, and Kalpana Shankar. 2019. "Talking about Metadata Labor: Social Science Data Archives, Professional Data Librarians, and the Founding of IASSIST." Pp. 83–113 in *Historical Studies in Computing, Information, and Society: Insights from the Flatiron Lectures*, *History of Computing*, edited by W. Aspray. New York: Springer International.
Durant, Jennie L. 2020. "Ignorance Loops: How Non-Knowledge about Bee-Toxic Agrochemicals Is Iteratively Produced." *Social Studies of Science* 50 (5): 751–77.
Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, Mass.: MIT Press.
———. 2017. "Knowledge Infrastructures for the Anthropocene." *Anthropocene Review* 4 (1): 34–43.
———. 2019. "Knowledge Infrastructures under Siege." Pp. 21–42 in *Data Politics*, edited by Didier Bigo, Engin Isin, and Evelyn Ruppert. New York: Routledge.
Edwards, Paul N., Steven J. Jackson, Melissa K. Chalmers, Geoffrey C. Bowker, et al. 2013. "Knowledge Infrastructures: Intellectual Frameworks and Research Challenges." Working paper. University of Michigan. http://hdl.handle.net/2027.42/97552
Feenberg, Daniel. R., and James M. Poterba. 1993. "Income Inequality and the Incomes of Very High Income Taxpayers: Evidence from Tax Returns." Pp. 145–77 in *Tax Policy and the Economy*, vol. 7. Cambridge, Mass.: National Bureau of Economic Research.
———.2000. "The Income and Tax Share of Very High Income Households, 1960–1995." *American Economic Review* 90 (2): 264–70.
Fisher, Jonathan, David Johnson, Jonathan Latner, Timothy Smeeding, and Jeffrey Thompson. 2018. "Estimating the Marginal Propensity to Consume Using the Distributions of Income, Consumption, and Wealth." Working paper. Washington Center for Equitable Growth. https://equitablegrowth.org/working-papers/marginal-propensity-consume/

Fixler, Dennis, and David Johnson. 2012. "Accounting for the Distribution of Income in the U.S. National Accounts." Paper presented at NBER Conference on Research in Income and Wealth. September 30.
Frickel, Scott, Sahra Gibbon, Jeff Howard, Joanna Kempner, Gwen Ottinger, and David J. Hess. 2010. "Undone Science: Charting Social Movement and Civil Society Challenges to Research Agenda Setting." *Science, Technology and Human Values* 35 (4): 444–73.
Frickel, Scott, and Michelle Edwards. 2014. "Untangling Ignorance in Environmental Risk Assessment." Pp. 215–33 in *Powerless Science? Science and Politics in a Toxic World*, edited by Soraya Boudia and Nathalie Jas. New York: Berghan Books.
Fourcade, Marion. 2009. *Economists and Societies*. Princeton, N.J.: Princeton University Press.
Fujimura, Joan H. 1987. "Constructing 'Do-Able' Problems in Cancer Research: Articulating Alignment." *Social Studies of Science* 17 (2): 257–93.
Gordon, Robert J. 1988. "U.S. Inflation, Labor's Share, and the Natural Rate of Unemployment." NBER Working Paper 2585. National Bureau of Economic Research. Cambridge, Mass.
Gould-Wartofsky, Michael. 2014. "The Invention of the '99 Percent': New Evidence on the Origins and Development of a Political Imaginary." Paper presented at the meetings of the American Sociological Society. San Francisco, August 16–19.
Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2021. "Tax Evasion at the Top of the Income Distribution: Theory and Evidence." NBER Working Paper 28542. National Bureau of Economic Research, Cambridge, Mass.
Hacker, Jacob S, and Paul Pierson. 2010. "Winner-Take-All Politics: Public Policy, Political Organization, and the Precipitous Rise of Top Incomes in the United States." *Politics and Society* 38 (2): 1–52.
Hacking, Ian. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.
Hall, Brian J., and Jeffrey B. Liebman. 1998. "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics* 113 (3): 653–91.
Hallett, Tim, Orla Stapleton, and Michael Sauder. 2019. "Public Ideas: Their Varieties and Careers." *American Sociological Review* 84 (3): 545–76.
Hirschman, Daniel. 2016*a*. "Inventing the Economy, or How We Learned to Stop Worrying and Love the GDP." Ph.D. Thesis. University of Michigan, Department of Sociology.
———. 2016*b* "Stylized Facts in the Social Sciences." *Sociological Science* 3:604–26.
———. 2021. "'Controlling for What?' Folk Economics, Legal Consciousness, and the Gender Wage Gap in the United States." SocArXiv. November 12. Doi:10.31235/osf.io/j8pra.
Huyser, Kimberly R., Arthur Sakamoto, and Isao Takei. 2010. "The Persistence of Racial Disadvantage: The Socioeconomic Attainments of Single-Race and Multi-Race Native Americans." *Population Research and Policy Review* 29 (4): 541–68.
Igo, Sarah E. 2007. *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. Cambridge, Mass.: Harvard University Press.
JEC (Joint Economic Committee). 1949. "Report on the Economic Report of the President, January 1949." Senate Report No. 88. 81st Congress, 1st Session. Washington, D.C.: Government Printing Office.
Jones, Carolyn C. 1988. "Class Tax to Mass Tax: The Role of Propaganda in the Expansion of the Income Tax during World War II." *Buffalo Law Review* 37:685.
Kaldor, Nicholas. 1961. *Capital Accumulation and Economic Growth*. New York: Macmillan.
Kaplan, Steven N., and Joshua Rauh. 2007. "Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?" NBER Working Paper 13270. National Bureau of Economic Research, Cambridge, Mass.
Katz, Lawrence F., and David H. Autor. 1999. "Changes in the wage structure and earnings inequality." Pp. 1463–1555 in *Handbook of Labor Economics*, Volume 3, Part A.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest, and Money*. London: Palgrave Macmillan.
Kopczuk, Wojciech, and Eric Zwick. 2020. "Business Incomes at the Top." *Journal of Economic Perspectives* 34 (4): 27–51.
Krueger, Alan. 1999. "Measuring Labor's Share." NBER Working Paper 7006. National Bureau of Economic Research, Cambridge, Mass.
Krugman, Paul. 1990. *Age of Diminished Expectations*. Cambridge, Mass.: MIT Press.
———. 1995. "Growing World Trade: Causes and Consequences." *Brookings Papers on Economic Activity* 1:327–77.
———. 1998. "Sharp Tongue, Maverick Mind." Interview with Orna Feldman. *Soundings: School of Humanities and Social Sciences at MIT*. Fall 1998. Soundings: Fall 98 (mit.edu) or http://www.pkarchive.org/others/krugmit.html.
———. 2002. "For Richer." *New York Times Magazine*, October 20.
Kuznets, Simon. 1934. *National Income, 1929–32*. Letter from the Acting Secretary of Commerce in Response to Senate Resolution 220. Senate Document No. 124. 73rd Congress, 2nd Session. Washington, D.C.: Government Printing Office.
———. 1953. *Shares of Upper Income Groups in Income and Savings*. Publications of the National Bureau of Economic Research No. 55. Cambridge, Mass.: NBER.
Lacey, Jim. 2011. *Keep from All Thoughtful Men: How U.S. Economists Won World War II*. Annapolis, Md.: Naval Institute Press.
Laumann, Edward O., Robert T. Michael, and John H. Gagnon. 1994. "A Political History of the National Sex Survey of Adults." *Family Planning Perspectives* 26 (1): 34–38.
Lebergott, Stanley. 1971. "Better Bricks for OBE's Structures." *Survey of Current Business* 51 (7, pt. 2): 117–8.
Leicht, Kevin T. 2008. "Broken Down by Race and Gender? Sociological Explanations of New Sources of Earnings Inequality." *Annual Review of Sociology* 34 (1): 237–55.
Mankiw, N. Gregory. 2003. *Macroeconomics*. 5th ed. Worth Publishers.
McCall, Leslie. 2013. *The Undeserving Rich: American Beliefs about Inequality, Opportunity, and Redistribution*. Cambridge: Cambridge University Press.
McGoey, Linsey. 2017. "The Elusive Rentier Rich: Piketty's Data Battles and the Power of Absent Evidence." *Science, Technology, and Human Values* 42 (2): 257–79.
McGonagle, Katherine A., Robert F. Schoeni, Narayan Sastry, and Vicki A. Freedman. 2012. "The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research." *Longitudinal and Life Course Studies* 3 (2): 268–84.
McKay, Dwanna L., Kirsten Vinyeta, and Kari Marie Norgaard. 2020. "Theorizing Race and Settler Colonialism within U.S. Sociology." *Sociology Compass*.
McNulty, Paul J. 1980. *The Origins and Development of Labor Economics*. Cambridge, Mass.: MIT Press.
Metcalf, Charles E. 1969. "The Size Distribution of Personal Income during the Business Cycle." *American Economic Review* 59 (4): 657–68.
Mincer, Jacob. 1958. "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy* 66 (4): 281–302.
———. 1997. "The Production of Human Capital and the Life Cycle of Earnings: Variations on a Theme." *Journal of Labor Economics* 15 (1, pt. 2): S26–S47. doi: 10.1086/209855.
Mirowski, Philip. 1984. "Physics and the Marginalist Revolution."' *Cambridge Journal of Economics* 8 (4): 361–79.
Morris, Martina, and Bruce Western. 1999. "Inequality in Earnings at the Close of the Twentieth Century." *Annual Review of Sociology* 25:623–57.
Mügge, Daniel, and Lukas Linsi. 2021. "The National Accounting Paradox: How Statistical Norms Corrode International Economic Data:" *European Journal of International Relations* 27 (2): 403–72.
NBER (National Bureau of Economic Research). 1921. *Income in the United States, Its Amount and Distribution, 1909–1919*. New York: Harcourt, Brace.

O'Connor, Alice. 2002. *Poverty Knowledge: Social Science, Social Policy, and the Poor in Twentieth-Century U.S. History*. Princeton, N.J.: Princeton University Press.

Peterson, David. 2017. "The Depth of Fields: Managing Focus in the Epistemic Subcultures of Mind and Brain Science." *Social Studies of Science* 47 (1): 53–74.

Pettit, Becky. 2012. *Invisible Men: Mass Incarceration and the Myth of Black Progress*. New York: Russell Sage Foundation.

Piketty, Thomas, and Emmanuel Saez. 2001. "Income Inequality in the United States, 1913–1998." Working Paper 8467. National Bureau of Economic Research, Cambridge, Mass.

———. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118 (1): 1–41.

Piketty, T., and E. Saez. 2006. "Response to: 'The top 1% . . . of What?' By Alan Reynolds." Manuscript. http://eml.berkeley.edu/~saez/answer-WSJreynolds.pdf

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Belknap Press.

———. 2020. *Capital and Ideology*. Cambridge, Mass.: Harvard University Press.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United States." *Quarterly Journal of Economics* 133 (2): 553–609.

Plantin, Jean-Christophe. 2019. "Data Cleaners for Pristine Datasets: Visibility and Invisibility of Data Processors in Social Science." *Science, Technology, and Human Values* 44 (1): 52–73.

Pullen, John. 2009. *The Marginal Productivity Theory of Distribution: A Critical History*. London: Routledge.

Quinn, Sarah. 2017. "'The Miracles of Bookkeeping': How Budget Politics Link Fiscal Policies and Financial Markets." *American Journal of Sociology* 123 (1): 48–85.

Radner, Daniel, and John Hinrichs. 1974. "Size Distribution of Income in 1964, 1970, and 1971." Pp. 19–31 in *Survey of Current Business*, October.

Reynolds, A. 2007. "Has U.S. Income Inequality Really Increased?" Policy Analysis 586. *Cato Institute*.

Rheinberger, Hans-Jörg. 1994. "Experimental Systems: Historiality, Narration, and Deconstruction." *Science in Context* 7 (1): 65–81.

Ribes, David, and Jessica Beth Polk. 2015. "Organizing for Ontological Change: The Kernel of an AIDS Research Infrastructure." *Social Studies of Science* 45 (2): 214–41.

Ricardo, David. (1817) 1996. *Principles of Political Economy and Taxation*. Amherst, N.Y.: Prometheus Books.

Rutherford, Malcolm. 2011. *The Institutionalist Movement in American Economics, 1918–1947: Science and Social Control*. Cambridge: Cambridge University Press.

Rutherford, R. S. G. 1955. "Income Distributions: A New Model." *Econometrica* 23 (3): 277–94.

Saez, Emmanuel, and Gabriel Zucman. 2020. "The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts." *Journal of Economic Perspectives* 34 (4): 3–26.

Sanders, Bernie. 2012. *The Speech: A Historic Filibuster on Corporate Greed and the Decline of Our Middle Class*. New York: Nation Books.

Sandmo, Agnar. 2014. "The Principal Problem in Political Economy: Income Distribution in the History of Economic Thought." 2:3–62 in *Handbook of Income Distribution*. Amsterdam: Elsevier Science Direct.

Schultz, T. Paul. 1964. "The Distribution of Personal Income: A Study of Statistics on the Size Distribution of Personal Income in the United States." Study for the JEC. 88th Congress, 2nd Session. Washington, D.C.: Government Printing Office.

Schultz, Theodore W. 1950. "Reflections on Poverty within Agriculture." *Journal of Political Economy* 58 (1):1–15.

Shankar, Kalpana, Kristin R. Eschenfelder, and Greg Downey. 2016. "Studying the History of Social Science Data Archives as Knowledge Infrastructure." *Science and Technology Studies* 29 (2): 63–72.

Smith, David M., and Gary J. Gates. 2001. "Gay and Lesbian Families in the United States: Same-Sex Unmarried Partner Households: A Preliminary Analysis of 2000 United States Census Data." Report. Human Rights Campaign, Washington, D.C.

Smith-Doerr, Laurel, Sharla Alegria, Kaye Husbands Fealing, Debra Fitzpatrick, and Donald Tomaskovic-Devey. 2019. "Gender Pay Gaps in U.S. Federal Science Agencies: An Organizational Approach." *American Journal of Sociology* 125 (2): 534–76.

Snipp, C. Matthew. 1992. "Sociological Perspectives on American Indians." *Annual Review of Sociology* 18:351–71.

Solovey, Mark. 2020. *Social Science for What? Battles over Public Funding for the "Other Sciences" at the National Science Foundation*. Cambridge, Mass.: MIT Press.

Stafford, Frank. 1986. "Forestalling the Demise of Empirical Economics: The Role of Microdata in Labor Economics Research." Pp. 1:387–423 in *Handbook of Labor Economics*, edited by Orley C. Ashenfelter and Richard Layard. Amsterdam: Elsevier.

Stapleford, Thomas A. 2009. *The Cost of Living in America: A Political History of Economic Statistics, 1880–2000*. Cambridge: Cambridge University Press.

Stone, Richard, and W. M. Stone. 1938. "The Marginal Propensity to Consume and the Multiplier: A Statistical Investigation." *Review of Economic Studies* 6 (1): 1–24.

Teixeira, Pedro N. 2007. *Jacob Mincer: The Founding Father of Modern Labor Economics*. Oxford: Oxford University Press.

Thomas, J. J. 1992. "Income Distribution and the Estimation of the Consumption Function: A Historical Analysis of the Early Arguments." *History of Political Economy* 24 (1): 153–81.

Tily, Geoff. 2009. "John Maynard Keynes and the Development of National Accounts in Britain, 1895–1941." *Review of Income and Wealth* 55 (2): 331–59.

Tomaskovic-Devey, Donald, and Dustin Avent-Holt. 2019. *Relational Inequalities: An Organizational Approach*. New York: Oxford University Press.

Walter, Maggie, and Chris Andersen. 2013. *Indigenous Statistics: A Quantitative Research Methodology*. London: Routledge.

Wold, H. O. A., and P. Whittle. 1957. "A Model Explaining the Pareto Distribution of Wealth." *Econometrica* 25 (4): 591–95.

Wolff, Edward N. 1997. *Economics of Poverty, Inequality and Discrimination*. Boston: South-Western College Publishing.

———. 2008. *Poverty and Income Distribution*, 2nd ed. Malden, Mass.: Wiley-Blackwell.

Yonay, Yuval P. 1998. *The Struggle over the Soul of Economics*. Princeton, N.J.: Princeton University Press.