The Journal of Economic Perspectives

A journal of the
American Economic Association

Winter 2025

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@aeapubs.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed at Sheridan Press, Hanover, Pennsylvania, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Winter 2025, Vol. 39, No. 1. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.

POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the USA.

The Journal of
Economic Perspectives

Contents

Volume 39 • Number 1 • Winter 2025

Statement of Purpose

The *Journal of Economic Perspectives* aims to bridge the gap between the general interest business and financial press and standard academic journals of economics. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 3–28

The 2023 Merger Guidelines and the Arc of Antitrust History

Daniel Francis

Which mergers are illegal? Today, despite more than a century of merger law, that question is surprisingly hard to answer. The main statute, Section 7 of the Clayton Act of 1914, tells us that a transaction is illegal if its effect "may be substantially to lessen competition."[1] But that language raises many questions. Does a merger "lessen competition" simply by reducing the number of rivals? By increasing prices, or profit margins? Is consumer harm necessary? And what about mergers that reduce prices: do they *increase* competition?

As times have changed, so too have views about the best way to give meaning to the statutory "competition" test. Speaking very broadly, over the last 60 years or so we can discern a broad secular trend in the merger work of courts and enforcers: from a "structuralist" perspective that measured harm to competition primarily by reference to increased market concentration, to a "welfarist" perspective that did so primarily by reference to higher prices and equivalent harms. Thus, in the 1960s, courts and antitrust agencies generally focused on whether a transaction increased market concentration, even blocking deals that promised to benefit consumers. By the 1980s, courts and enforcers generally focused on whether a transaction would harm consumer welfare, usually through higher prices or lower output. This reflected a broader resettlement of the entire antitrust project on a welfarist foundation.

■ *Daniel Francis is Assistant Professor of Law, New York University School of Law, New York, New York. From 2018–2021, he served successively as Senior Counsel, Associate Director for Digital Markets, and Deputy Director, Bureau of Competition, Federal Trade Commission, Washington, DC. His email address is daniel.francis@law.nyu.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20241415.

[1] It also prohibits transactions of which the effect may be to "tend to create a monopoly," but enforcers and courts have generally not focused on that language.

The turn to welfare was broadly welcomed by many enforcers and scholars as a journey toward enlightened policy. But about a decade ago, an emerging school of "Neo-Brandeisian" or "antimonopoly" critics charged that antitrust's welfarist turn had betrayed its original mission, and demanded radical reforms. Following the election of President Joseph Biden in 2020, leaders of this group were appointed to head the antitrust agencies: the Antitrust Division of the Department of Justice (DOJ) and the Federal Trade Commission (FTC).

Among the new leadership's highest priorities was a revision of the agencies' merger guidelines. These guidelines—first published in 1968 and periodically updated—declare federal merger enforcement policy, but they do not create binding law. Courts, not the enforcers, are the authoritative interpreters of the Clayton Act. However, the guidelines matter in at least two ways. First, when interpreting the law, courts have often found the guidelines persuasive "in view of the many years of thoughtful analysis they represent" (*United States v. Anthem, Inc.*, 855 F.3d 345, 349 [D.C. Cir. 2017]). Innovations in the guidelines have often led to significant changes in court practice (Greene 2006). Second, the guidelines explain the agencies' own policy, including what deals and theories will be investigated and litigated. As such, they cast a long deterrent shadow.

This essay offers a first draft of the place of the 2023 Merger Guidelines in antitrust history. I first trace the long arc of modern US merger policy, with an emphasis on the broad shift from industry structure to consumer welfare. I then outline the Neo-Brandeisian effort to reform merger law, leading to the 2023 Merger Guidelines at the intersection of the two narratives. I centrally argue that the new guidelines represent a strategically ambiguous challenge to welfare's dominance in modern antitrust. In some important respects, the text can plausibly be understood *either* as a welfarist document *or* as an endorsement of non-welfarist policies that would, in at least some cases, sacrifice welfare in the pursuit of other goals. Perhaps more than any previous iteration, the 2023 guidelines leave a reader uncertain of what ultimate criterion, or set of criteria, will guide enforcement. This ambiguity comes with costs—including confusion for those who rely on the guidance, like businesses and courts—but it may have helped the agencies to navigate a complicated political moment. In effect, ambiguity may have been a feature, not a bug.

I will offer two further perspectives on the document from two different imagined futures. From the perspective of a post-revolutionary future in which welfarist antitrust has been displaced by some other paradigm, I will suggest that the 2023 guidelines will have made an important, but solely deconstructive, contribution to that revolution, by undermining welfarism without articulating an alternative paradigm. Alternatively, from the perspective of a post-evolutionary future in which US antitrust policy remains welfarist, I will suggest that the 2023 guidelines may one day be seen as an incremental development within that very same tradition. From this future, historians may look back at the storm of controversy that has swirled around the document, and wonder what all the fuss was about. In the meantime, it exemplifies the ambitions and limits of an effort to bring about a generational reform of merger policy.

The Arc of Merger History

The Merger Statute

The foundational US merger statute is Section 7 of the Clayton Act of 1914.[2] In its modern form, it applies to all acquisitions of shares or assets, including "horizontal" transactions (between actual or potential rivals, such as two coffee-store chains) as well as "vertical" ones (between parties at different levels of a supply chain, or suppliers of complements, such as a coffee-store chain and a coffee bean distributor).

Mergers often have multiple effects. Horizontal mergers reduce the number of rivals and may create some pricing power by eliminating a substitute, or by facilitating coordination among the remaining players, but they may also result in cost savings that exert downward pressure on prices. Vertical mergers may reduce transaction costs and create incentives to lower prices, but they may also enable the merged firm to foreclose rivals' access to inputs, distribution, customers, or complements, or to coordinate with rivals.

Section 7 states that a transaction is unlawful if its "effect . . . may be substantially to lessen competition, or to tend to create a monopoly." And the central problem is that it is not obvious what "competition" is supposed to mean, nor what it means to "lessen" it (Francis 2024; Kaplow 2024).[3] Competition has many dimensions, and economists have never come up with an "index of rivalry" to integrate them (Kaplow 2024). This creates a puzzle when mergers have mixed effects. Many transactions reduce the headcount of rivals and increase market concentration (that is, the extent to which a market is dominated by a few players), or improve the relative strength of the merged firm against its rivals, but also benefit consumers, including through lower prices.[4] How does a "competition" test apply to such mergers?

The 1914 Congress that wrote Section 7 did not pin down the meaning of "competition," and at least some of them knew they were leaving a terrible puzzle for courts.[5] Senator John W. Weeks of Massachusetts (51 Cong. Rec. 15,988, Oct. 1, 1914) despaired of the bill that would become the Clayton Act:

> I have been listening to the debate on this subject off and on for several weeks, and I have come to the conclusion that there is not a Member of the Senate who really knows what the result of this legislation is going to be, either from

[2] The US antitrust system also includes, among other things, prohibitions on agreements in restraint of trade and monopolization, in Sections 1 and 2 of the Sherman Act, respectively.

[3] There are plenty of other puzzles too, like the meaning of "may be," and of "tend to create a monopoly"—the latter long neglected by courts and enforcers, as noted above—but I will leave those for another day.

[4] Commentators disagree on how common merger efficiencies are in practice, and the evidence is ambiguous (for example, see Rose and Sallet 2020; Ohlhausen and Owings 2023; Shapiro 2024).

[5] The 1950 amendments, noted below, did not alter the "substantially . . . lessen competition" standard: as a result, the 1914 legislative debates, not the 1950 ones, seem apposite.

> a legal standpoint or from the standpoint of the great industrial operations of this country.

As it turned out, Section 7 initially fell flat as a result of some unfortunate drafting. For example, acquisitions of assets, rather than stock, fell outside the original 1914 text. But in 1950 Congress amended the statute to fix the problems, and throughout the 1960s the Supreme Court issued a stream of Section 7 decisions.

In the main, these 1960s cases took a "structuralist" approach, driven by robust skepticism of mergers that increased market concentration, with a resulting emphasis on market shares and entry barriers. The Court was particularly suspicious of deals in markets that were trending toward increased concentration, regardless of the reasons for the trend, and of whether consumers were suffering or benefiting as a result. During this period the Court adopted a rule that a merger could be presumed illegal based on structural information alone (the "structural presumption").

This structuralist approach was heavily influenced by the "structure-conduct-performance" research paradigm in economic scholarship. This paradigm, grounded in the work of Edward Mason, Joe Bain, Leonard Weiss, and others, reflected a view that concentrated market structures protected by barriers to entry tended to lead to harmful conduct, including tacit or explicit collusion, and thus generally to worse market performance. Scholars in the structure-conduct-performance school argued that, as conduct was difficult to detect, enforcement and policy efforts should generally focus on avoiding concentrated markets (Bain 1968; Weiss 1979; Panhans 2024).

The influence of economic scholarship on the structural turn in the law is very clear. As early as 1953, Alfred Kahn (1953) had discerned a current of academic advocacy for a structuralist approach. A decade later, the Supreme Court itself cited a string of prominent economists—Joe Bain, Betty Bock, Carl Kaysen, Fritz Machlup, Edward Mason, Donald Turner, and George Stigler—in its blockbuster decision articulating the structural presumption, *United States v. Philadelphia National Bank* (374 US 321 [1963]). The law clerk who wrote the opinion, Richard Posner, would go on to become a leader of the "law and economics" movement in the American legal academy, which drew on economic scholarship to guide law reform (Posner 2015). More generally, the influence of the structure-conduct-performance view was clear across the antitrust enterprise. The Justice Department chose structure-conduct-performance leader Leonard Weiss as an expert witness for its flagship antitrust case against IBM, with much of the expert battle in that case located within the structure-conduct-performance paradigm (Panhans 2024). Donald Turner, who coauthored a prominent structure-conduct-performance text, even led the Antitrust Division of the US Department of Justice, where he inaugurated the practice of bringing economists directly into Division leadership (Williamson 2002).

The Court's version of merger structuralism condemned transactions that were unlikely to harm anyone except less efficient rivals, and which may well have benefited consumers. In *Brown Shoe Co. v. United States* (370 US 294 [1962]), for example, the Court condemned a merger between two vertically integrated shoe manufacturer-retailers that would have led to modest market shares, fretting that "[i]f a merger achieving 5% control were now approved, we might be required to approve future merger efforts by . . . competitors seeking similar market shares." And in *United States v. Von's Grocery Co.* (384 US 270, 278 [1966]), it condemned a merger that created a firm with a 7.5 percent share of a competitive Los Angeles grocery market.

The Court of this era showed little interest in whether a concentrated market remained competitive enough to preclude harm (for example, see *United States v. Von's Grocery Co.*, 384 US 270, 278 [1966]). It proscribed vertical deals that would give the merged firm an "advantage" over rivals, without much interest in whether consumers would be better or worse off as a result (for example, *United States v. E.I. du Pont de Nemours & Co.*, 353 US 586, 607 [1957]). Most strikingly of all, it did not seem interested in arguments that a merger would be good for consumers—for example, by lowering costs and thus prices (for example, *FTC v. Procter & Gamble Co.*, 386 US 568, 580 [1967]). After all, such efficiencies would tend to increase the market share and competitive strength of the merged firm, fueling further concentration.

This viewpoint reflected an explicit distance from welfarism, which the Court attributed to Congress's vision: "Congress was aware that some mergers which lessen competition may also result in economies," the Court shrugged in *Procter & Gamble*, "but it struck the balance in favor of protecting competition." In *Brown Shoe*, the goal was explicitly structural: "Congress appreciated that occasional higher costs and prices might result from the maintenance of fragmented industries and markets. It resolved these competing considerations in favor of decentralization."

In this climate, the US Department of Justice issued its first merger guidelines in 1968. These guidelines stated that "the primary role of Section 7 enforcement is to preserve and promote market structures conducive to competition," and so "[m]arket structure is the focus of the Department's merger policy." Following the Supreme Court's lead, the guidelines expressed a strong "structural presumption" that increased concentration leads to harm, and that enforcement decisions would primarily turn on the market shares of the parties. Only in "exceptional circumstances" would other factors be relevant. They stated that in an *unconcentrated* market, a merger of two firms each with a 5 percent share would "ordinarily" be challenged. Mergers were more likely to be illegal in the presence of a "significant trend toward increased concentration." On the ground that efficiencies could not easily be measured and were unlikely to outweigh harms, they would generally be ignored.

The 1968 Merger Guidelines declared that large vertical mergers "usually" resulted in harms "wholly disproportionate" to benefits. Thus, the merger of an upstream firm with a 10 percent share of sales and a downstream firm with a 6 percent share of purchases would "ordinarily" be challenged. And the guidance expressed particular concern with deals that might "entrench" market power, including through "increas[ing] product differentiation," regardless of consumer harm.

The Antitrust Revolution

The 1970s saw a genuine antitrust revolution. During this decade, a turn to effects-based consumer welfarism—the view that antitrust analysis should focus on whether a specific practice or transaction harmed or benefited consumers, particularly through effects on price, output, and quality—can be seen across the antitrust landscape, including but not limited to merger law. This revolution had many causes, including changes in scholarly views, in the national political and economic climate, and in Supreme Court and other government personnel (Director and Levi 1956; Bork 1978; Werden 1992; Berman 2022; Francis and Sprigman 2024).

Changes in the academy, in particular, were an important factor. By this time, the structure-conduct-performance paradigm was losing support among academic economists, including because of the insight that concentrated markets might be the result of superior performance, not a cause of inferior performance (Demsetz 1973; Panhans 2024).

A package of analytical methods and background views associated with the "Chicago School" was rising in scholarly popularity, just as the political winds were turning in ways that favored its adherents and their prescriptions. This package included, for example, the propositions that policy should aim to maximize some measure of aggregate economic welfare; that basic universalizing price theory was a surer guide to antitrust policy than cross-sectional empirical work of the kind favored by structure-conduct-performance scholars; that concentration as such was no cause for concern; that market power was usually self-correcting, including because entry barriers were rare; and that government intervention was often more costly than beneficial (Bork 1978; R.A. Posner 1979; Hovenkamp and Scott Morton 2020).

As the 1970s unfolded, the Court—with five new members joining between 1969 and 1975—decisively embraced a welfarist approach across its non-merger antitrust docket. In horizontal-agreement law, for example, the Court steered away from an earlier formalistic antipathy to coordination, embracing effects-based welfarism instead. In *BMI, Inc. v. CBS, Inc.* (441 US 1 [1979]), a decision permitting a joint licensing arrangement, the Court even said that some conduct involving "'price fixing' in the literal sense" was lawful. In vertical-agreement law, welfarism supplanted long-standing concerns with trader freedom and independence. In *Continental T.V., Inc. v. GTE Sylvania Inc.* (433 US 36 [1977]), the Court reshaped the analysis of vertical distribution restraints, centering consumer impact and disclaiming interest in dealer freedom. In 1979, the Court explicitly stated that Congress designed the Sherman Act, the main antitrust statute, as a "consumer welfare prescription" (*Reiter v. Sonotone Corp.*, 442 US 330 [1979]).

Consistent with this trend, in the early 1970s the Supreme Court started to row back from the strong structuralism of its 1960s merger cases. In *United States v. General Dynamics* (415 US 486 [1974]), the Court permitted a merger between two coal producers in a fairly concentrated market, emphasizing evidence that one of the parties faced an "unpromising" future. *General Dynamics* did not reject structuralism, but it signaled interest in case-specific facts and rebuttals. This was consistent with the shift toward welfarism, which implied attention to matters beyond structure, including efficiencies as well as the expected future behavior of the merged firm.

But the Court's merger docket was soon cut short. In 1974, Congress substantially repealed legislation that had created a fast procedural track to the Supreme Court—bypassing the intermediate appellate courts—for civil antitrust cases brought by the US Department of Justice. This fast track had been critical for cases involving mergers, which are typically time-sensitive. After its repeal, few merging parties could litigate up to the Supreme Court. Thus, after 1975, the Court would return to Section 7 just a handful of times, and never for a merits adjudication. So the old structuralist decisions remained "good law" even as their conceptual foundations were widely abandoned, and even as the Court explicitly embraced welfarism across the rest of its antitrust docket.

After the Revolution

New merger guidelines were issued in 1982, reflecting the priorities of the new Reagan administration, the fall of structuralism in the academy, and a changed judicial climate. Among other things, structural arguments were no longer landing with courts in antitrust cases as they used to (Scheffman, Coate, and Silvia 2003).

The 1982 guidance was immediately recognized as a repudiation of high structuralism (Fox 1982; Schwartz 1983). Gone was the statement that the "primary role" of merger enforcement is to preserve deconcentrated markets. In its place was a new declaration: "The unifying theme of the Guidelines is that mergers should not be permitted to create or enhance 'market power' or to facilitate its exercise." Market power was defined as "[the] ability . . . profitably to maintain prices above competitive levels for a significant period of time." And a new concern was recognized: "the Department seeks to avoid unnecessary interference with that larger universe of mergers that are either competitively beneficial or neutral."

Structure was still important. The 1982 revision reframed the structural presumption, expressing it in terms of the Herfindahl-Hirschman Index (HHI).[6] The guidance affirmed that "[a]dditional concentration resulting from mergers [in highly concentrated markets] is a matter of significant competitive concern." The Department would "likely" challenge mergers with a post-merger HHI of 1,800 or more that increased the HHI by 100 points or more. This was a significant relaxation. For example, given a pre-merger market with 20 firms, each with a 5 percent share, a merger of any two would "ordinarily" have been challenged under the 1968 guidelines. But such a deal, raising the HHI from 500 to 550, would be far below the 1982 thresholds. The new approach was soon embraced by courts (Greene 2006).

But the role of structure in the 1982 Merger Guidelines was qualified by increasing concern with price harms. Market definition was explicitly tied to the prospect of power over price, with the introduction of a "hypothetical monopolist test"

[6] The HHI is the sum of the squares of the market shares of the participants in a market. Thus, a market with two participants, each with a 50 percent market share, has an HHI of $50^2 + 50^2 = 5{,}000$.

that defined antitrust markets by reference to a threat of price increases (Werden 1992).[7] Trends toward concentration—a ground of special concern in the 1968 text and in the Supreme Court's output—were ignored. Likelihood of post-merger entry was identified as a reason to permit a merger, if it meant that "existing competitors could not succeed in raising price for any significant period of time . . ." The "entrenchment" theories were cut out entirely.

Crucially, the 1982 guidelines explicitly presented a concentration increase as a basis to presume welfare harms, subject to rebuttal, not as a matter of concern in its own right. Structure's relevance was grounded in the inference that "[o]ther things being equal, [structure] affects the likelihood that one firm, or a small group of firms, could successfully exercise market power" or coordinate. The new text stated that "even in concentrated markets, it is desirable to allow firms some scope for merger activity in order to achieve economies of scale and to permit exit from the market." But in the 1982 guidelines, potential efficiencies from a merger were still practically irrelevant: "Except in extraordinary cases, the Department *will not consider a claim of specific efficiencies as a mitigating factor.*"

In 1984, the merger guidelines were further amended, marking another step away from strong structuralism. The revisions downgraded structure to a mere "starting point for analyzing . . . competitive impact," indicating that, in every case, "[b]efore determining whether to challenge a merger, the Department will consider all other relevant factors that pertain to its competitive impact." It emphasized the relationship between efficiencies and "lower prices to consumers," and said for the first time that efficiencies would weigh against a decision to take enforcement action. The 1984 text also turned the concern about trends toward concentration on its head, noting that widespread vertical integration could weigh in *favor* of a merger because it could suggest that such integration yielded "substantial economies."

In 1986, the Supreme Court sent a strong signal that it, too, had left its 1960s jurisprudence behind. In *Cargill v. Monfort of Colorado* (479 US 104 [1986]), two beef-packing companies proposed to merge, creating the second-largest firm in the market. A rival sued—channeling the spirit of the Court's own earlier cases—on the ground that the deal would increase market concentration and, through merger efficiencies, result in low prices that would drive the rival out. But the Court rejected that theory outright. What the plaintiff alleged, the Court protested, "is not activity forbidden by the antitrust laws. It is simply . . . vigorous competition. To hold that the antitrust laws protect competitors from the loss of profits due to such price competition would, in effect, render illegal any decision by a firm to cut prices in order to increase market share. The antitrust laws require no such perverse result." *Cargill*

[7] The hypothetical monopolist test tests a candidate market definition by asking whether a profit-maximizing monopolist of the candidate set of products would impose a small but significant price increase on at least one product.

did not explicitly disown *Brown Shoe* or its cousins, but in substance and orientation it sharply rebuked the Court's earlier cases. Lower courts got the message.[8]

New horizontal merger guidelines in 1992 and 1997 continued the trend. The 1992 guidance demonstrated that a shift toward price effects need not always favor defendants, by articulating a "unilateral effects" theory of harm. This theory of harm, flowing from economic scholarship in the 1980s, is grounded not in contribution to market-wide concentration and coordination, but in the creation of pricing power from the acquisition of a close substitute, giving the merged firm the "unilateral" power to raise prices (Baker 2003). The 1992 document also introduced a detailed treatment of post-merger entry, providing a framework to assess when entry would protect against price increases.

The 1997 revisions added specific criteria for an efficiencies rebuttal of competitive concerns. These revisions came close to saying explicitly that what makes a merger "anticompetitive" is consumer harm (emphasis added):

> *The Agency will not challenge a merger if cognizable efficiencies are of a character and magnitude such that the merger is not likely to be anticompetitive in any relevant market.* To make the requisite determination, the Agency considers whether cognizable efficiencies likely would be sufficient to *reverse the merger's potential to harm consumers* in the relevant market, e.g., by preventing price increases in that market.

This text strongly implies a "price standard." If prices will go up, the merger is anticompetitive; if they will go down, it is procompetitive (Coate and Heimert 2009). A bottom line of this kind was consistent with the practice of the Supreme Court, and lower courts too, in non-merger cases, where a focus on welfare impacts was now routine. From this perspective, of course, there is no clear reason to treat price effects resulting from efficiencies any differently from other effects of the merger on price.

The 2010 revisions to the merger guidelines took another step toward a comprehensive consumer welfarism. Since 1982, the guidelines had declared that their "unifying theme" was that "mergers should not be permitted to create or enhance market power or to facilitate its exercise." The 2010 revision clarified that the concern was deals that "encourage one or more firms to *raise price, reduce output,*

[8] For example, in 1986, the Seventh Circuit praised the Federal Trade Commission for "studiously avoid[ing]" the Court's 1960s cases, noting that they were of doubtful force, and for instead "inquir[ing] into the probability of harm to consumers" (*Hospital Corp. of America v. FTC*, 807 F.2d 1381 [7th Cir. 1986]). In 1990, repeatedly citing the 1984 Guidelines, the DC Circuit held that "[e]vidence of market concentration simply provides a convenient starting point for a broader inquiry into future competitiveness," and stated that "[a]lthough the Supreme Court has not overruled [its 1960s] section 7 precedents, it has cut them back sharply" (*United States v. Baker Hughes Inc.*, 908 F.2d 981 [D.C. Cir. 1990]). In 1991, the Eleventh Circuit held—citing the 1984 Guidelines—that efficiencies could rebut a prima facie case of harm to competition, if they were "significant" and "ultimately would benefit competition and, hence, consumers" (*FTC v. University Health, Inc.*, 938 F.2d 1206 [11th Cir. 1991]).

diminish innovation, or otherwise harm customers as a result of diminished competitive constraints or incentives." Concentration was now firmly secondary: "[t]he measurement of market shares and market concentration is not an end in itself, but is useful to the extent it illuminates the merger's likely competitive effects."[9] The 2010 guidelines also substantially expanded the treatment of unilateral effects. And they raised for the first time the HHI thresholds at which the agencies would infer harm, as in practice enforcers had focused on transactions involving higher concentration levels than the extant thresholds suggested (Hovenkamp and Shapiro 2018; Rose and Shapiro 2022).

By 2010, the welfarist consensus was, in most important respects, long settled in professional antitrust circles. The prevailing view: mergers that increase price (or otherwise harm consumers) are bad; those that decrease it are good. Agency practice, likewise, was firmly welfarist in its treatment of harms and benefits alike (Coate and Heimert 2009; Baer 2015; Feinstein 2017; Hoffman 2018; Shapiro 2010). The 2020 Vertical Merger Guidelines, which later supplemented the 2010 document, expressed the same basic view.

Courts, too, were strongly welfarist in merger cases. Courts typically focused on predicting effects, sometimes expressed skepticism that structural evidence *alone* could fully discharge a plaintiff's burden of proof, and usually counted efficiencies as a point in favor of legality (for example, *St. Alphonsus Med. Ctr.-Nampa Inc. v. St. Luke's Health Sys., Ltd.*, 778 F.3d 775 [9th Cir. 2015]; *FTC v. Penn State Hershey Med. Ctr.*, 838 F.3d 327 [3d Cir. 2016]; *FTC v. Staples, Inc.*, 190 F. Supp. 3d 100 [D.D.C. 2016]). To be sure, sometimes a court would rely on structure alone, or nearly alone, to infer the illegality of a deal (as in *FTC v. H.J. Heinz Co.*, 246 F.3d 708 [D.C. Cir. 2001]), and sometimes a court would doubt an "efficiencies defense," although usually evaluating it anyway (for example, see *United States v. Anthem, Inc.*, 855 F.3d 345 [D.C. Cir. 2017]). But the differences were generally of degree, not kind.

The Neo-Brandeisian Movement and the Road to the 2023 Guidelines

The Neo-Brandeisian (or "antimonopoly") movement, which emerged about a decade ago, is a coalition of writers and policy advocates who favor deep legal reforms to address concerns about economic power and its consequences—particularly those relating to inequality and exploitation.[10] The movement's name recalls former Supreme Court Justice Louis Brandeis, and some of his views are emphasized by the school (Crane 2019).

[9] Indeed, as Kaplow points out in this symposium, the term HHI is mentioned just once in the unilateral effects and coordinated effects sections of the 2010 guidelines, and then only to downplay its importance compared to "the value of diverted sales."

[10] Representative contributions include, for example, Khan (2017), Khan and Vaheesan (2017), Wu (2018b, 2019), Vaheesan (2019a, b), OMI (2019), Teachout (2019), and Hanley (2024).

Neo-Brandeisian perspectives on merger policy often include (1) the view that merger enforcement has been much too permissive since antitrust's welfarist turn; (2) proposals to restore the primacy of the Supreme Court's structuralist decisions, and more generally of "law" over "economics"; (3) a preference for simpler, structural rules rather than case-specific effects analysis; and (4) a belief that merger efficiencies are rare in practice and should not count in favor of the legality of a merger as a matter of law (for examples, see Wu 2018b; Wu 2019; OMI 2019; Vaheesan 2019b; Glick et al. 2023; Hanley 2024).

Following the 2020 election, President Biden appointed several Neo-Brandeisians to antitrust leadership roles, including Lina Khan as Chair of the Federal Trade Commission, Jonathan Kanter as head of the Deprtment of Justice Antitrust Division, and Tim Wu as Special Assistant to the President.

These leaders strongly criticized the policy and doctrine of the last "four decades": the era of welfarist antitrust. Kanter (2022a) declared that "the era of lax enforcement is over, and the new era of vigorous and effective antitrust law enforcement has begun." Khan (2022a, 2023b) promised to reverse "decades of lax . . . enforcement" and "broad government inaction." Wu (2023) described an effort to "turn the battleship" and to "reverse" and "reset" a "40-year trend."

They particularly criticized the dominance of consumer welfare in antitrust doctrine. Wu (2018a) argued that, while consumer harm "might matter in some contexts," "to say it always matters, and is indeed the lodestone of the law, is . . . unsupportable and can . . . border on the ridiculous." Other antimonopoly writers expressed similar criticisms (Vaheesan 2019a; Teachout 2019; Steinbaum and Stucke 2020; Bush 2023). Neo-Brandeisian scholars instead favored a competitive process standard that would reflect a broader, though not fully specified, set of values.

A revision of the merger guidelines was a "key project" in the antitrust reset (Khan 2021b). "The 1982 merger guidelines," Khan (2023a) stated, "set forth assumptions and frameworks that have stayed intact and shaped merger policy for 40 years . . . . As federal enforcers, our job is to enforce the laws as Congress has written and as the courts have interpreted." The existing guidelines, the leaders indicated, were "outdated" and led to under-enforcement (FTC 2022; Khan 2021b, 2022a; Kanter 2022c).

Officials made several specific criticisms. First, they argued that existing merger guidance improperly disregarded "the law," and particularly the Supreme Court cases of the 1960s (Khan 2022b; Khan 2023a; Lawrence 2023). Second, they criticized the centrality in merger review of measuring and evaluating welfare effects, like impacts on price and output. Kanter (2022b) argued that antitrust enforcers should not "attempt to weigh impacts," stating: "We cannot predict specific outcomes in the future and we should not try, nor is that what the law requires." He contrasted "competition" not with "consumer harm," but with "concentration" (Kanter 2022b). Third, they criticized the agencies' receptivity to efficiency arguments. Khan (2022b, 2021a) explicitly rejected "the notion that efficiencies might justify a merger that was otherwise illegal," citing "clear precedent rejecting such a balancing test given the text of the statute," and lamented the turn away from structure toward effects.

The 2023 Draft Merger Guidelines

In July 2023, the Federal Trade Commission and Department of Justice published a draft revision of the merger guidelines. The draft provided an initial expression of the agencies' vision, and provides important context for an assessment of the final document.

The draft was a complete rewrite of the merger guidance. Some of the changes were consistent with welfarist enforcement. For example, the structural presumption thresholds were returned to their lower pre-2010 levels. Also, new discussions of ways in which transactions might cause harm were added, including material relating to potential competition, vertical mergers, platforms, buy-side competition, and labor markets. Welfarist scholars had previously proposed similar additions (for example, Salop and Scott Morton 2021; Shapiro 2021; Salop 2022; Francis 2022; Rose and Shapiro 2022).

But other changes departed from welfarism, and condemned some transactions that would benefit, or not harm, consumers. At least five features were salient. First, the draft jettisoned the 40-year-old orienting language that referred to market power and (in the 2010 text) to welfare harms. In its place was an undefined "competition" test, which echoed statutory text but was not identified with any particular definition or metric.

Second, while the 2010 Horizontal Merger Guidelines had focused on methodologies for economic analysis, the 2023 draft instead centered 13 individual "Guidelines," mostly in the form "Mergers Should Not . . . ." These injunctions condemned categories of mergers without regard to their welfare impacts. For example, the draft stated that mergers should not "significantly increase concentration in highly concentrated markets," "eliminate a potential entrant in a concentrated market," or "entrench or extend a dominant position," apparently regardless of whether they harmed or benefited consumers. Similarly, the draft invoked the 1962 *Brown Shoe* case to support a concern that a deal might "restrict options along the supply chain, depriving rivals of a fair opportunity to compete." It stated that furthering a trend toward concentration or vertical integration was a reason to condemn a merger, regardless of consumer impact, and that a merger could create a "conflict of interest" by uniting a platform with a trading partner.

Third, the draft repeatedly cited the Supreme Court's 1960s structuralist cases, disfavoring more recent welfarist authorities.

Fourth, while the draft acknowledged the relevance of efficiencies, it emphasized earlier Supreme Court language that "possible economies [from a merger] cannot be used as a defense to illegality." And it stated that "efficiencies are not cognizable if they will accelerate a trend toward concentration . . . or vertical integration."

Fifth, the draft significantly weakened the relationship between market definition—antitrust's tool for measuring the bounds of competition—and the threat of price harms, instead promoting a qualitative, intuitive standard.[11]

[11] For a full discussion of the draft's relationship with market definition, see Kaplow's essay in this symposium.

The Reaction to the Draft

The publication of the draft triggered a wide-ranging public conversation. The Federal Trade Commission and the US Department of Justice hosted a series of workshops, and officials engaged with commentators and the public at events across the country. Where the 2010 revisions of the merger guidelines had attracted only a handful of comments,[12] the 2023 draft attracted more than 3,000, including many from the lay public.

Those professionally concerned with antitrust—including scholars, policy advocates, practitioners, and market participants—expressed a variety of views. Some criticized the draft for failing to banish welfarism entirely and called for more aggressive structural rules and the exclusion of efficiencies (for example, OMI et al. 2023; Bush, Glick, and Lozada 2023). Others criticized the revision root and branch, often arguing that the draft ignored modern law and expressed an anti-merger, anti-business policy (for example, US Chamber of Commerce 2023; CEI 2023; CAGW 2023).

Comments from moderates often took an intermediate view, supporting more aggressive guidance but objecting to the break with welfarism. Specific concerns in this vein included (1) the risk of social harm from a merger policy that targeted concentration, rather than higher prices and other harms; (2) the dangers of misrepresenting existing law by "cherry picking" old structuralist cases to the exclusion of more recent ones; (3) the draft's hostility to vertical integration and transactions that contribute to dominance, regardless of impact on consumers; (4) the draft's hostility to concentration trends, regardless of consumer impact; (5) the draft's skeptical and vague treatment of merger efficiencies; and (6) the draft's demotion of a price-based approach to market definition (for various perspectives, see Baker et al. 2023; Crane 2023; Francis 2023a; Hovenkamp 2023a; Koegel, McDowall, and Michelozzi 2023; Salop 2023; Scott Morton 2023; Shapiro 2023).

Above all, the post-draft debate centered on the draft's relationship with welfarism. The draft had sketched a merger policy that, to a significant extent, would diverge from modern welfarist antitrust. Some commenters demanded that the agencies go further and entirely reject the welfarist paradigm; others demanded a clear recommitment to it. The agencies were at a crossroads.

Rorschach's Merger Guidelines

In December 2023, the antitrust agencies released the final 2023 Merger Guidelines. That document was structured around eleven individual "guidelines," presented in Table 1, with additional sections addressing rebuttal arguments and analytical tools.

[12] The comments are no longer available on the FTC website, but they are archived on the Wayback Machine at https://web.archive.org/web/20121012054545/https:/www.ftc.gov/os/comments/hmgrevisedguides/index.

Table 1
The Individual "Guidelines" in the 2023 Merger Guidelines

Guideline 1: Mergers raise a presumption of illegality when they significantly increase concentration in a highly concentrated market.

Guideline 2: Mergers can violate the law when they eliminate substantial competition between firms.

Guideline 3: Mergers can violate the law when they increase the risk of coordination.

Guideline 4: Mergers can violate the law when they eliminate a potential entrant in a concentrated market.

Guideline 5: Mergers can violate the law when they create a firm that may limit access to products or services that its rivals use to compete.

Guideline 6: Mergers can violate the law when they entrench or extend a dominant position.

Guideline 7: When an industry undergoes a trend toward consolidation, the agencies consider whether it increases the risk a merger may substantially lessen competition or tend to create a monopoly.

Guideline 8: When a merger is part of a series of multiple acquisitions, the agencies may examine the whole series.

Guideline 9: When a merger involves a multi-sided platform, the agencies examine competition between platforms, on a platform, or to displace a platform.

Guideline 10: When a merger involves competing buyers, the agencies examine whether it may substantially lessen competition for workers, creators, suppliers, or other providers.

Guideline 11: When an acquisition involves partial ownership or minority interests, the agencies examine its impact on competition.

Source: Available from the US Department of Justice at https://www.justice.gov/atr/2023-merger-guidelines.

On my reading, the final document effectively declines to make the choice between welfarism and non-welfarism that commenters had demanded. Instead, it opts for ambiguity: it can plausibly bear both a welfarist *and* a non-welfarist reading. Like a Rorschach test, the meaning of the 2023 Merger Guidelines depends upon what one expects, hopes, or fears to find there. Above all, it invites, but does not answer, a basic question: *Is a tendency to harm consumers (or other trading partners) necessary for condemnation under the 2023 guidelines?*

The closest thing in the final text to an explicit commitment to welfarism is the document's orienting statement. It reads:

> Section 7 of the Clayton Act . . . prohibits mergers and acquisitions where "in any line of commerce or in any activity affecting commerce in any section of the country, the effect of such acquisition may be substantially to lessen

> competition, or to tend to create a monopoly." Competition is a process of rivalry that incentivizes businesses to offer lower prices, improve wages and working conditions, enhance quality and resiliency, innovate, and expand choice, among many other benefits. Mergers that substantially lessen competition or tend to create a monopoly increase, extend, or entrench market power and deprive the public of these benefits. Mergers can lessen competition when they diminish competitive constraints, reduce the number or attractiveness of alternatives available to trading partners, or reduce the intensity with which market participants compete.

This passage is intricately crafted. It does *not* say: "A merger lessens competition for the purposes of these Guidelines when it causes, or tends to cause, harm to consumers or other trading partners by increasing or entrenching market power." (This would paraphrase the 2010 guidelines.) Instead, it indicates that welfare harms are significant, without specifying their precise relationship with merger policy or the guidelines. The third sentence states that unlawful mergers "deprive the public of . . . benefits," which suggests that illegality involves some kind of harmful impact, but it does not focus on traditional consumer welfare harms. Moreover, the fourth sentence locates harm to competition in factors like "dimin[ution of] competitive constraints," loss of alternatives as such, and lost "intensity" of competition. "Competition" remains the dominant, undefined standard. And while a tendency to reduce consumer welfare is *one* way in which it might be lessened, the passage leaves room for the view that this is not the *only* way.[13]

To be sure, one could play close-reading games like this with any document, including any previous version of the guidelines. But the document's text and context suggest that some distance from welfarism is a deliberate and important feature of the text, not a quibble or trick of the light.

Above all, many of the document's individual guidelines retain an ambiguous relationship with welfare harms, including because the individual guidelines make little or no reference to such harms. For the most part, the eleven "guidelines" roughly correspond to those in the non-welfarist draft, with the peremptory "mergers should not" formulation replaced by possibility language: "mergers can violate the law when . . . ." This phrasing creates a puzzle for transactions that implicate the central concern of a guideline but do not, in a particular case, seem to involve a tendency to inflict welfare harms. This would include, for example, a merger that is unlikely to result in consumer harm but involves one party with a 30 percent market share and will mildly increase market concentration (implicating Guideline 1); eliminates "substantial competition between" parties (implicating

[13] Among other things, the principle of decision is unspecified. Should an analyst test for welfare harms, and label that exercise a "harm to competition" test? Or should the analyst test for something else, and label that exercise a measure of tendency to inflict harms?

Guideline 2);[14] eliminates a potential entrant into a "concentrated" market (implicating Guideline 4); provides some "structural" reason for concern, despite the apparent absence of ability or incentive to foreclose rivals (implicating Guideline 5); contributes to market power or may result in power affecting competition in another market (implicating Guideline 6); furthers a trend toward concentration or vertical integration (implicating Guideline 7); or combines a platform owner and platform participant (implicating Guideline 9). The document also includes a catchall section, warning that other mergers might harm "competition" on grounds facially unconnected to welfare harms, including "due to the structure of the acquisition or the acquirer."

Moreover, the document pointedly embraces structuralist Supreme Court cases that famously have little or nothing to do with welfarism. The 1962 *Brown Shoe* in particular is cited repeatedly, including for its theory of competitive harm, and the text prominently quotes the statement that "possible economies [from a merger] cannot be used as a defense to illegality" (*FTC v. Procter & Gamble Co.*, 386 US 568, 580 [1967]).

The document's context implies the same thing. The Neo-Brandeisian leaders who announced and oversaw the revision had expressly stated that the revision was motivated in part by a desire to shift merger policy away from the modern (that is, welfarist) paradigm. Khan had explicitly stated that efficiencies should not, as a matter of law, weigh in favor of a merger. Kanter had stated that merger analysis should not turn on the prediction of future welfare effects. Officials had repeatedly emphasized a desire to foreground "the law" and to reject the approach of the last "four decades." In context, it is hard to understand this as anything other than a declared turn toward structuralist Supreme Court cases, and away from the welfarism that marginalized them.

Moreover, the orienting statement itself was effectively retrofitted to a draft that was itself evidently not a welfarist text, and which otherwise underwent only modest changes, leaving individual guidelines with little or no reference to the prospect of price increases or equivalent harms. Indeed, if the orienting statement itself were redacted, a reader would probably not infer the existence of a welfarist commitment from reading the rest of the text.

Finally, the draft-and-comment process had put a very bright spotlight on the guidelines' relationship with welfarism. The issue cannot have been anything other than top of mind.

Nevertheless, it is *also* true that the final version of the guidelines clearly marks a step back toward welfare from the open non-welfarism of the draft. The orienting statement quoted above does not clearly commit to welfare harm as a limiting

[14] The 2023 unilateral-effects guideline indicates that the magnitude of the concern is a function of the substantiality of pre-merger competition between the firms, and it downplays the importance of resulting pricing power or other harms. By contrast, the 2010 text stated that "a merger between firms selling differentiated products may diminish competition by enabling the merged firm to profit by *unilaterally raising the price of one or both products* above the pre-merger level" (emphasis added).

principle, but it does emphasize its importance. The final document also states that "[m]arket concentration and the change in concentration due to the merger are often useful indicators of a merger's risk of substantially lessening competition," thus inserting a clear conceptual gap between concentration and the document's conception of "competition." Overall, the final guidelines are materially less hostile to efficiencies than the draft had been.

So the final 2023 guidelines take a step back toward welfare, but they retain much that is not welfarist, and they leave plenty of room for confusion about the relationship between welfare and the undefined "competition" standard that orients the document. The Congressional Research Service (2024) captures it neatly:

> While [the final version of the Merger Guidelines] references market power, it preserves the . . . draft's primary emphasis on "competition." According to the new language, increased market power can result from mergers that substantially lessen "competition," but the focus remains on "competition" as an independent concept. . . . [S]everal parts of the finalized Guidelines suggest that the agencies regard market structure as a central consideration in analyzing possible harm to "competition."

After decades of increasingly explicit welfarism, this change is a big one. The 2023 guidelines imply that the agencies may challenge at least some mergers that benefit (or do not harm) consumers and other trading partners, and that they may invite courts to interpret the law to block such deals. This would, of course, involve a policy that will sometimes sacrifice consumer welfare—on some occasions perhaps even the welfare of all affected consumers (Kaplow and Shavell 2001). Moreover, even if the agencies remain resolutely welfarist in their choices about which investigations and litigations to pursue, the *risk* of non-welfarist enforcement may itself deter beneficial or harmless transactions. Indeed, several commentators have suggested that the Biden administration's policies are deterring lawful and desirable merger and acquisition activity (Wilson and Phillips 2021; Froeb, Tschantz, and Werden 2024; de Bodt et al. 2024; Shapiro 2024; Williams 2024). More generally, ambiguity impairs the document's ability to *guide*.

The step back from welfare may also present challenges in court. Previous guidelines have owed much of their persuasive force to judicial confidence that they represent elaborations of existing law rather than departures from it (Francis 2022; Hovenkamp 2024). This confidence may now be qualified in light of observations that, in some respects, the 2023 Guidelines "pursu[e] theories that are not reasonably calculated to identify mergers that truly harm competition" as traditionally understood (Hovenkamp 2024).

But this ambiguity may bring benefits as well as costs. The comment process had revealed that the draft's disconnection from welfarism was unacceptable to many pro-enforcement moderates, which created a problem. If a critical mass of expert opinion had condemned the guidelines as a partisan departure from settled law and sound policy, the project could have been in vain: courts would likely reject them, and a succeeding presidential administration would surely scrap them. But there was limited room to retreat: after the heavy emphasis laid on the "reset" and the break with "decades" of "failed" policy, a mere scrub-and-polish of the 2010 guidelines would have compromised credibility on a flagship Neo-Brandeisian effort.

The result was a recipe for a compromise text. The final document centers "competition": it states that welfare harms are important for an assessment of competition, but leaves room to conclude that they are not determinative and that other factors may require condemnation of a deal that would not be condemned on welfare grounds. Just as it had done in the formulation of the Clayton Act itself in 1914, the "competition" concept provided the basis for an "incompletely theorized agreement" among groups that could agree that they favored "competition," but could not converge on more abstract theories or more specific applications (for example, compare Sunstein 1995; Hills 2023). Each group could find its preferred policy within "competition," albeit at the cost of a clear standard (Francis 2024).

Early reactions to the final 2023 Merger Guidelines suggest that the compromise has been partly, but not fully, successful. The final version has generally received a warmer welcome than the draft. Some writers have hailed the document as a successful integration of rigorous economic methods with progressive, pro-enforcement goals (Salop 2024). Others have criticized it for failing to return to the Supreme Court's approach of the 1960s, including in its accommodation of efficiency arguments (Glick, Lande, and Bush 2024). And some moderate reformers have expressed a mixture of praise and criticism, welcoming the new material on theories of harm and the rebalanced thresholds but renewing objections to the document's retreat from the clear welfarism of the 2010 guidance and from modern judicial practice (Hovenkamp 2024; Shapiro 2024).

The ambiguity of the 2023 guidelines may also confer some tactical benefits. The final document is probably close *enough* to welfarist law and practice to enjoy a meaningful measure of judicial confidence. In turn, this may allow the agencies to invoke the document in support of incremental reforms, perhaps even on non-welfarist margins. Ultimately, the final text may be a better tool for reform than a more radical document—like the July 2023 draft—would have been.

But strategic ambiguity has its limits, and the document will only sustain divergent readings for so long. The dueling readings will clash in court, the agencies will have to take a position on the document's meaning—including in response to merging-party arguments that it is untethered from the law or relies on obsolete cases—and courts will write opinions that interpret and construct its "real" meaning. History, in effect, will pick a lane.

The Revolutionary Road: Vanguard, Not Manifesto

If today's Neo-Brandeisian moment ripens into a true antitrust revolution, the 2023 Merger Guidelines will surely come to be seen as a watershed. They will represent the turn of the welfarist tide; a step back from the preoccupation with welfare impacts, price, output, and so on; and a pivot back toward "competition," "law," and a wider perspective. From this future, the document's decisive contribution will be the reopening of antitrust's imagination.

But the contribution of the 2023 guidance to a post-welfarist revolution will be as vanguard, not manifesto. For, although it downgrades the role of welfare and price, the 2023 guidance does not offer a clear alternative criterion. While it emphasizes some concerns beyond the promotion of economic welfare—particularly concentration, market power, and "competition"—none of these seems able to serve as an orienting criterion for merger policy.

A return to a full-blown 1960s-style concentration standard does not seem plausible. In fact, as noted above, the final document makes affirmative efforts to disclaim that notion. A policy that aims at deconcentration as such—at the expense of higher prices—is likely to prove unacceptable to voters, Congress, and public-minded courts (Hovenkamp 2019). Most courts have not taken the structuralist cases seriously for a long time—including because the Supreme Court itself has made such a firm break from them in its own non-merger cases. Even Don Turner (1982), a primary author of the 1968 Guidelines, has conceded that there is "no defensible economic theory that would support such broad general prohibitions" as those in the 1968 text. The earlier structure-conduct-performance approach remains "discredited" in the academy (Berry, Gaynor, and Scott Morton 2019).

A literal focus on market power seems similarly implausible. A merger can augment market power without harming anyone except rivals—for example, by reducing costs and prices while increasing margins. A crusade against margins as such is hardly better politics, or policy, than one against concentration as such (but see Posner 2024).

That leaves the document's own lodestar: "competition." But, as I have argued elsewhere, the concept of promoting (or harming) "competition" or the "competitive process" is not self-executing in antitrust (Francis 2024; Kaplow 2024). For example, suppose that a merger will eliminate a potential entrant into a moderately concentrated market; or that it will promote a trend toward concentration or vertical integration; or that it will contribute to or complement a position of "dominance"; or that it will unite a platform owner with a platform participant. In these and other cases, the 2023 text does not clearly specify what test will determine whether that merger will be challenged.[15] Previous guidance, from 1968 to 2010, was consistently clearer about the core principle—whether concentration or welfare harm—that would distinguish good mergers from bad.

In this respect the 2023 guidelines crystallize a puzzle raised by much Neo-Brandeisian writing. Much of it does not clearly specify operationalizable non-welfarist criteria that could plausibly guide antitrust doctrine. To return to where we started: which mergers should be illegal? And why? Any true antitrust revolution will need clear, fresh answers to those questions.

[15] For a detailed discussion in the context of extension-of-dominance theories, see the essay by Shapiro in this symposium.

The Evolutionary Reading: Long Live the 1982 Guidelines

An antitrust revolution is not guaranteed. The Neo-Brandeisian moment may never ripen into a genuine revolution, especially given the results of the 2024 election; instead, antitrust may just continue to develop in a broadly welfarist direction.

At present, this seems to me the most likely outcome, regardless of whether the 2023 guidelines are retained, withdrawn, or amended. As noted above, no plausible alternative to welfarism has yet been articulated, and the old objections to structuralism have not been overcome (Hovenkamp 2023b). In addition, half a century of adjudication has generated a robust edifice of welfarist doctrine that is not remotely controversial among courts. Outside the courtroom, non-welfarist antitrust reforms have not attracted strong support.

To be sure, the Neo-Brandeisian movement has had genuine successes. It has triggered a broad conversation about antitrust and competition, surfaced or resurfaced some submerged or marginalized concerns, and punctured the false appearance of post-political neutrality that had settled over antitrust. These achievements are real, even if they have served in large part to invigorate and rebalance welfarist antitrust rather than to replace it (Francis 2023b).

But the Biden administration's antitrust enforcement successes have, in the main, been welfarist, not Neo-Brandeisian: motivated by welfare harms and pleaded and proved accordingly (for a summary of recent activity, see Nylen 2023). Indeed, many flagship enforcement programs have built on the efforts of previous presidential administrations. This includes, for example, efforts under the Trump administration to challenge harmful conduct by Big Tech platforms (FTC 2019; Complaint, *FTC v. Facebook, Inc.*, No. 1:20-cv-3590 [D.D.C. Dec. 9, 2020]; Complaint, *United States v. Google LLC*, No. 1:20-cv-03010 [D.D.C. Oct. 20, 2020]); to protect labor competition (Simons 2020; FTC 2020; Mekki 2019); to consider rulemaking to address labor non-competes (FTC 2020); and to target harmful vertical mergers (*United States v. AT&T, Inc.*, 916 F.3d 1029 [D.C. Cir. 2019]; Hoffman 2018). This is the stuff of evolutionary welfarism, not of Neo-Brandeisian revolution.

From a welfarist future, it may be hard even to remember that the document ever seemed revolutionary. Virtually every sentence in the 2023 Guidelines could be interpreted sensibly subject to a proviso that a tendency to inflict welfare harm is necessary for liability. Indeed, on that reading the 2023 guidelines are a cousin of the 2010 horizontal guidance and the 2020 vertical guidance, with modest amendments fleshing out particular paths to harm that were less trodden without being entirely novel. From this perspective, for example, the "structural" inference of vertical merger illegality is a familiar inference of market power from market shares; the "entrenchment and extension" theories do not innovate on existing merger law; a trend toward concentration is just a background fact framing the future of a market; a "conflict of interest" just means an incentive to foreclose; the downplaying of price encourages greater attention to non-price harms; and so on. From this view, the orienting statement should be understood as a commitment to welfarism.

Of course, as noted above, the 2023 guidelines contain sentences and ideas that, if read *without* the tie to welfarism, could reach much further. But the same has been true, to at least some extent, of every set of merger guidelines, including the 2010 guidelines. Most previous guidelines, for example, contain the statement that the "unifying theme" of merger control is that mergers should not be permitted to create market power or facilitate its exercise. But no one thought that the 2010 guidelines expressed hostility to market power as such. Harm was the touchstone in practice, whatever the text said.

The point is that we always read our merger guidelines with at least one eye on the agencies that have issued them, and the rhetoric and practice that accompany them. When agency leaders criticize welfarism and deny that efficiencies should weigh in favor of a merger's legality, when they promise a radical break from decades of "lax" and "failed" policy, and when the final text is less than perfectly clear, it should not be a surprise that readers wonder how much radicalism they should find in the text. But if the 2023 guidelines had been issued at a different moment, with different background music but with similar text, it may never have occurred to anyone to think of them as anything other than a welfarist text. We may yet find that—once the background music has changed and the heat of the moment has receded—that that very same reading becomes the most natural one. We may even forget that any other reading was ever plausible.

Conclusion

Almost 40 years ago, Eleanor Fox (1987) described "a battle for the soul of antitrust": a fundamental conflict over the identity and direction of the entire enterprise. Today, that conflict has broken out anew. The 2023 Merger Guidelines largely avoid picking a side—itself a revolutionary choice after decades of clear welfarist hegemony—but by centering the purported "competition" criterion, and by raising some tough questions about what it is supposed to mean, they may invite and frame a valuable conversation. Ultimately, the fate of the Neo-Brandeisian challenge to the arc of merger law will be decided by enforcers (including, in the first instance, the appointees of the incoming Trump administration), courts, and the voting public. And only in retrospect will the winners of that struggle be able to claim the 2023 guidelines as their own.

■ *Thanks to Krisha Cerilli, Scott Hemphill, Louis Kaplow, David Lawrence, Rohan Pai, Nina Pavcnik, Carl Shapiro, Timothy Taylor, and Heidi Williams for helpful comments and conversations.*

References

Baer, William. 2015. "Integration, Consolidation, Competition in the Wake of the Affordable Care Act." Speech, The New Health Care Industry Conference, Yale University, November 13, 2015. https://www.justice.gov/opa/speech/assistant-attorney-general-bill-baer-delivers-remarks-new-health-care-industry-conference.

Bain, Joe S. 1968. *Industrial Organization*. New York: Wiley.

Baker, Jonathan B. 2003. "Why Did the Antitrust Agencies Embrace Unilateral Effects." *George Mason Law Review* 12 (1): 31–38.

Baker, Jonathan B., Andrew I. Gavil, Richard J. Gilbert, Herbert Hovenkamp, Michael L. Katz, A. Douglas Melamed, Fiona Scott Morton, Daniel L. Rubinfield, Carl Shapiro, and Howard Shelanski. 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1406.

Berman, Elizabeth Popp. 2022. *Thinking Like an Economist: How Efficiency Replaced Equality in U.S. Public Policy*. Princeton: Princeton University Press.

Berry, Steven, Martin Gaynor, and Fiona Scott Morton. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives* 33 (3): 44–68.

de Bodt, Eric, Jean-Gabriel Cousin, Micah S. Officer, and Richard Roll. 2024. "The Deterrence Effect of M&A Regulatory Enforcements." *Journal of Corporate Finance* 85: 102559.

Bork, Robert H. 1978. *The Antitrust Paradox: A Policy at War with Itself*. New York: Basic Books.

Brozen, Yale. 1969. "Significance of Profit Data for Antitrust Policy." *Antitrust Bulletin* 14 (1): 119–40.

Bush, Darren. 2023. "Antitrust Heretics and the Problem of One-Sided Bias: An Essay." http://dx.doi.org/10.2139/ssrn.4601891.

Bush, Darren, Mark Glick, and Gabriel Lozada. 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1411.

Citizens Against Government Waste (CAGW). 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1371.

Coate, Malcolm B., and Andrew J. Heimert. 2009. "Merger Efficiencies at the Federal Trade Commission 1997–2007." FTC Economic Issues Working Paper.

Competitive Enterprise Institute (CEI). 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1415.

Congressional Research Service (CRS). 2024. "The 2023 Merger Guidelines: 30 Analysis and Issues for Congress." CRS Report LSB11138.

Crane, Daniel A. 2019. "How Much Brandeis Do the Neo-Brandeisians Want?" *Antitrust Bulletin* 64 (4): 531–39.

Crane, Daniel. 2023. "Law over Economics in the 2023 Draft Merger Guidelines." *Network Law Review*, August 2. https://www.networklawreview.org/cranes-cartel-four/.

Demsetz, Harold. 1973. "Industry Structure, Market Rivalry, and Public Policy." *Journal of Law and Economics* 16 (1): 1–9.

Director, Aaron, and Edward H. Levi. 1956. "Law and the Future: Trade Regulation." *Northwestern University Law Review* 51 (2): 281–96.

Federal Trade Commission (FTC). 2019. "FTC's Bureau of Competition Launches Task Force to Monitor Technology Markets." Press release, February 26. https://www.ftc.gov/news-events/news/press-releases/2019/02/ftcs-bureau-competition-launches-task-force-monitor-technology-markets.

FTC. 2020. "Non-Competes in the Workplace: Examining Antitrust and Consumer Protection Issues." Press release, January 9. https://www.ftc.gov/news-events/events/2020/01/non-competes-workplace-examining-antitrust-consumer-protection-issues.

FTC. 2022. "Oversight of the Enforcement of the Antitrust Laws." Speech, United States Senate Subcommittee on Competition Policy, Antitrust, and Consumer Rights of the US Senate Committee on the Judiciary, Washington, DC, September 20, 2022. https://www.ftc.gov/legal-library/browse/oversight-enforcement-antitrust-laws-prepared-statement-federal-trade-commission-subcommittee.

Feinstein, Deborah L. 2017. "To Know Where You're Going, Look at Where You've Been." Remarks of February 22, 2017. https://www.ftc.gov/news-events/news/speeches/know-where-youre-going-look-where-youve-been.

Fox, Eleanor M. 1982. "The New Merger Guidelines—A Blueprint for Microeconomic Analysis." *Antitrust*

Bulletin 27 (3): 519–91.

Fox, Eleanor M. 1987. "The Battle for the Soul of Antitrust." *California Law Review* 75 (3): 917–23.

Francis, Daniel. 2022. "Revisiting the Merger Guidelines: Protecting an Enforcement Asset." *CPI Antitrust Chronicle*: 1–14.

Francis, Daniel. 2023a. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1358.

Francis, Daniel. 2023b. "Reflections on the Revolution in Antitrust." *Journal of Antitrust Enforcement* 11 (2):185–90.

Francis, Daniel. 2024. "Antitrust without Competition." *Duke Law Journal* 74: 353–439.

Francis, Daniel, and Christopher Jon Sprigman. 2024. *Antitrust: Principles, Cases, and Materials*. 2nd ed. Chicago: American Bar Association.

Froeb, Luke M., Steven Tschantz, and Gregory J. Werden. 2024. "Deterrence in Merger Review: Likely Effects of Recent U.S. Policy Changes." *CPI Antitrust Chronicle*. https://www.pymnts.com/cpi-posts/deterrence-in-merger-review-likely-effects-of-recent-u-s-policy-changes/.

Glick, Mark, Robert H. Lande, and Darren Bush. 2024. "The Efficiency Rebuttal in the New Merger Guidelines: Bad Law and Bad Economics." *Antitrust* 38 (3): 20–26.

Glick, Mark, Gabriel A. Lozada, Pavitra Govindan, and Darren Bush. 2023. "The Horizontal Merger Efficiency Fallacy," Institute for New Economic Thinking Working Paper 212.

Greene, Hillary. 2006. "Guideline Institutionalization: The Role of Merger Guidelines in Antitrust Discourse." *William and Mary Law Review* 48 (3): 771–857.

Hanley, Daniel. 2024. "Redefining the Relevant Market: Abandonment or Return to *Brown Shoe*." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4404081.

Hills, Roderick M., Jr. 2023. "Strategic Ambiguity and Article VII: Why The Framers Decided Not to Decide." *Journal of American Constitutional History* 1 (3): 379–448.

Hoffman, Bruce. 2018. "Vertical Merger Enforcement at the FTC." Speech, Credit Suisse 2018 Washington Perspectives Conference, Washington, DC, January 10, 2018. https://www.ftc.gov/news-events/news/speeches/vertical-merger-enforcement-ftc.

Hovenkamp, Herbert. 2019. "Is Antitrust's Consumer Welfare Principle Imperiled?" *Journal of Corporation Law* 45 (1): 65–94.

Hovenkamp, Herbert. 2023a. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1280.

Hovenkamp, Herbert. 2023b. "*Brown Shoe* Merger Policy and the Glorification of Waste." *CPI Antitrust Chronicle*: 1–7.

Hovenkamp, Herbert. 2024. "The 2023 Merger Guidelines: Law, Fact, and Method." *Review of Industrial Organization* 65 (1): 39–77.

Hovenkamp, Herbert, and Fiona Scott Morton. 2020. "Framing the Chicago School of Antitrust Analysis." *University of Pennsylvania Law Review* 168 (7): 1843–78.

Hovenkamp, Herbert, and Carl Shapiro. 2018. "Horizontal Mergers, Market Structure, and Burdens of Proof." *Yale Law Journal* 127 (7): 1996–2025.

Kahn, Alfred E. 1953. "Standards for Antitrust Policy." *Harvard Law Review* 67 (1): 28–54.

Kanter, Jonathan. 2022a. "Antitrust Enforcement: The Road to Recovery." Keynote Address, University of Chicago Stigler Center, Chicago, April 21, 2022. https://www.justice.gov/opa/speech/assistant-attorney-general-jonathan-kanter-delivers-keynote-university-chicago-stigler.

Kanter, Jonathan. 2022b. "Remarks at New York City Bar Association's Milton Handler Lecture." Speech, Milton Handler Lecture, New York, May 18, 2022. https://www.justice.gov/opa/speech/assistant-attorney-general-jonathan-kanter-delivers-remarks-new-york-city-bar-association.

Kanter, Jonathan. 2022c. "Respecting the Antitrust Laws and Reflecting Market Realities." Keynote Address, Georgetown Antitrust Symposium, Washington, DC, September 13, 2022. https://www.justice.gov/opa/speech/assistant-attorney-general-jonathan-kanter-delivers-keynote-speech-georgetown-antitrust.

Kaplow, Louis. 2024. *Rethinking Merger Analysis*. Cambridge, MA: MIT Press.

Kaplow, Louis, and Steven Shavell. 2001. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle." *Journal of Political Economy* 109 (2): 281–86.

Kaysen, Carl, and Donald F. Turner. 1965. *Antitrust Policy: An Economic and Legal Analysis*. Cambridge, MA: Harvard University Press.

Khan, Lina M. 2017. "Amazon's Antitrust Paradox." *Yale Law Journal* 126: 710–805.

Khan, Lina M. 2021a. "Remarks Regarding the Proposed Rescission of the FTC's Approval of the 2020

Vertical Merger Guidelines." FTC File P810034.

Khan, Lina M. 2021b. "Memorandum to Commission Staff and Commissioners Regarding Vision and Priorities for the FTC." FTC Memorandum, September 22, 2021. https://www.ftc.gov/legal-library/browse/cases-proceedings/public-statements/memo-chair-lina-m-khan-commission-staff-commissioners-regarding-vision-priorities-ftc.

Khan, Lina M. 2022a. "Keynote Remarks at the International Competition Network." Keynote Address, International Competition Network, Berlin, May 6, 2022. https://www.ftc.gov/system/files/ftc_gov/pdf/Remarks%20of%20Chair%20Lina%20M.%20Khan%20at%20the%20ICN%20Conference%20on%20May%206%2C%202022_final.pdf.

Khan, Lina M. 2022b. "Remarks at Fordham Annual Conference on International Antitrust Law and Policy." Speech, Fordham Annual Conference on International Antitrust Law and Policy, New York, September 16, 2022. https://www.ftc.gov/news-events/news/speeches/remarks-chair-lina-m-khan-prepared-delivery-fordham-annual-conference-international-antitrust-law.

Khan, Lina M. 2023a. "Remarks at Fordham 50th Annual Conference on International Antitrust Law and Policy." Speech, Fordham 50th Annual Conference on International Antitrust Law and Policy, September 22, 2023. https://www.ftc.gov/news-events/news/speeches/remarks-chair-lina-m-khan-prepared-delivery-fordham-50th-annual-conference-international-antitrust.

Khan, Lina M. 2023b. "Remarks at Stanford Institute for Economic Policy Research." Speech, Stanford Institute of Economic Policy Research, Stanford, CA, November 2, 2023. https://www.ftc.gov/news-events/news/speeches/remarks-chair-lina-m-khan-stanford-institute-economic-policy-research.

Khan, Lina, and Sandeep Vaheesan. 2017. "Market Power and Inequality: The Antitrust Counterrevolution and Its Discontents." *Harvard Law and Policy Review* 11: 235–94.

Koegel, Kate Maxwell, Ana McDowall, and Lorenzo Michelozzi. 2023. "A Summary of Comments on the 2023 Draft Merger Guidelines." *Antitrust Source*: 1–11.

Lawrence, David. 2023. "Remarks at the Georgetown Center for Business and Public Policy." Speech, Georgetown Center for Business and Public Policy, Washington, DC, July 26, 2023.

Mekki, Doha. 2019. *Statement before the Subcommittee on Antitrust, Competition Policy, and Consumer Rights, Committee on the Judiciary, US Senate for a Hearing Entitled "Antitrust and Economic Opportunity: Competition in Labor Markets."* October 29.

Nylen, Leah. 2023. "Biden Antitrust Enforcers Set New Record for Merger Challenges." *Bloomberg News*, December 18. https://news.bloomberglaw.com/mergers-and-acquisitions/biden-antitrust-enforcers-set-new-record-for-merger-challenges.

Ohlhausen, Maureen K., and Taylor M. Owings. 2023. "Evidence of Efficiencies in Consummated Mergers." US Chamber of Commerce Paper 20230601.

Open Markets Institute (OMI). 2019. "Restoring Antimonopoly through Bright-Line Rules." ProMarket, April 26. https://www.promarket.org/2019/04/26/restoring-antimonopoly-through-bright-line-rules/.

OMI et al. 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1502.

Panhans, Matthew T. 2024. "The Rise, Fall, and Legacy of the Structure-Conduct-Performance Paradigm." *Journal of the History of Economic Thought* 46 (3): 337–57.

Posner, Eric A. 2024. "Market Power, Not Consumer Welfare: A Return to the Foundations of Merger Law." *Antitrust Law Journal* 86 (1): 205–50.

Posner, Richard A. 1979. The Chicago School of Antitrust Analysis. *University of Pennsylvania Law Review* 127 (4): 925–48.

Posner, Richard A. 2015. "Philadelphia National Bank at 50: An Interview with Judge Richard Posner." Interview by C. Scott Hemphill. *Antitrust Law Journal* 80 (2): 205–17.

Rose, Nancy L., and Jonathan Sallet. 2020. "The Dichotomous Treatment of Efficiencies in Horizontal Mergers: Too Much? Too Little? Getting it Right." *University of Pennsylvania Law Review* 168: 1941–84.

Rose, Nancy L., and Carl Shapiro. 2022. "What Next for the Horizontal Merger Guidelines?" *Antitrust* 36 (2): 4–13.

Salop, Steven C. 2022. "A Suggested Revision of the 2020 Vertical Merger Guidelines." *Antitrust Bulletin* 67 (3): 371–89.

Salop, Steven C. 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1364.

Salop, Steven C. 2024. "The 2023 Merger Guidelines: A Post-Chicago and Neo-Brandeisian Integration." *Review of Industrial Organization* 65 (1): 79–128.

Salop, Steven C., and Fiona Scott Morton. 2021. "The 2010 HMGs Ten Years Later: Where Do We Go from Here?" *Review of Industrial Organization* 58 (1): 81–101.

Scheffman, David, Malcolm Coate, and Louis Silvia. 2003. "Twenty Years of Merger Guidelines Enforcement at the FTC: An Economic Perspective." *Antitrust Law Journal* 71 (1): 277–318.

Schwartz, Louis B. 1983. "The New Merger Guidelines: Guide to Governmental Discretion and Private Counseling or Propaganda for Revision of the Antitrust Laws?" *California Law Review* 71 (2): 575–603.

Scott Morton, Fiona. 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1437.

Shapiro, Carl. 2010. "The 2010 Horizontal Merger Guidelines: From Hedgehog to Fox in Forty Years." *Antitrust Law Journal* 77: 701–59.

Shapiro, Carl. 2021. "Antitrust: What Went Wrong and How to Fix It." *Antitrust* 35 (3): 33–45.

Shapiro, Carl. 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1393.

Shapiro, Carl. 2024. "Evolution of the Merger Guidelines: Is This Fox Too Clever by Half?" *Review of Industrial Organization* 65 (1): 147–75.

Simons, Joseph J. 2020. "Statement of the FTC Chairman Regarding Announcement that Aveanna Healthcare and Maxim Healthcare Services Have Terminated Their Acquisition Agreement." January 30. https://www.ftc.gov/news-events/news/press-releases/2020/01/statement-ftc-chairman-regarding-announcement-aveanna-healthcare-maxim-healthcare-services-have.

Steinbaum, Marshall, and Maurice E. Stucke. 2020. "The Effective Competition Standard." *University of Chicago Law Review* 87 (2): 595–623.

Sunstein, Cass R. 1995. "Incompletely Theorized Agreements." *Harvard Law Review* 108 (7): 1733–72.

Teachout, Zephyr. 2019. "Antitrust Law, Freedom, and Human Development." *Cardozo Law Review* 41 (3): 1081–1140.

Turner, Donald F. 1982. "Observations on the New Merger Guidelines and the 1968 Merger Guidelines." *Antitrust Law Journal* 51 (2): 307–15.

US Chamber of Commerce. 2023. Comment on "Draft Merger Guidelines for Public Comment." https://www.regulations.gov/comment/FTC-2023-0043-1440.

Vaheesan, Sandeep. 2019a. "The Profound Nonsense of Consumer Welfare Antitrust." *Antitrust Bulletin* 64 (4): 479–94.

Vaheesan, Sandeep. 2019b. "Two-and-a-Half Cheers for 1960s Merger Policy." *Harvard Law School Antitrust Association* (blog), December 12. https://orgs.law.harvard.edu/antitrust/2019/12/12/two-and-a-half-cheers-for-1960s-merger-policy/.

Weiss, Leonard W. 1979. "Structure-Conduct-Performance Paradigm and Antitrust." *University of Pennsylvania Law Review* 127 (4): 1104–40.

Werden, Gregory J. 1992. "The History of Antitrust Market Delineation." *Marquette Law Review* 76 (1): 123–215.

Williams, Roger. 2024. "Chairman Williams Writes to FTC, DOJ Regarding New Mergers and Acquisitions Guidelines." Press release, April 16. https://smallbusiness.house.gov/news/documentsingle.aspx?DocumentID=405947.

Williamson, Oliver E. 2002. "The Merger Guidelines of the U.S. Department of Justice—In Perspective." https://www.justice.gov/archives/atr/merger-guidelines-us-department-justice-perspective.

Wilson, Christine S., and Noah Joshua Phillips. 2021. "Dissenting Statement of Commissioners Christine S. Wilson and Noah Joshua Phillips Regarding the Statement of the Commission on Use of Prior Approval Provisions in Merger Orders." Speech, Federal Trade Commission, Washington, DC, October 29. https://www.ftc.gov/system/files/documents/public_statements/1598095/wilson_phillips_prior_approval_dissenting_statement_102921.pdf.

Wu, Timothy. 2018a. "The 'Protection of the Competitive Process' Standard." Columbia Public Law Research Paper 14-612.

Wu, Timothy. 2018b. *The Curse of Bigness: Antitrust in the New Gilded Age.* New York: Columbia Global Reports.

Wu, Timothy. 2019. "The Utah Statement: Reviving Antimonopoly Traditions for the Era of Big Tech." *Medium*, November 18. https://onezero.medium.com/the-utah-statement-reviving-antimonopoly-traditions-for-the-era-of-big-tech-e6be198012d7.

Wu, Timothy. 2023. "A New Era: The Biden Administration's Antitrust Policy." Milton Handler Lecture, New York City Bar Association, May 25. https://documents.nycbar.org/files/WU_Milton_Handler_Lecture.pdf.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 29–52

Improving Economic Analysis in Merger Guidelines

Louis Kaplow

Mergers and acquisitions shape the economic landscape by changing which assets and activities are collected within various firms. This activity is central to the dynamic evolution of a market economy in moving assets from low- to high-value uses, facilitating innovation, and enabling creative destruction. But anticompetitive mergers can destroy social value by reducing the attractiveness of merged firms' offerings to customers and suppliers and by otherwise inhibiting efficient resource flows.

Merger guidelines that originated in the United States have spread throughout the developed world. Modern merger guidelines articulate how competition agencies balance potential harms and benefits in an important regulatory setting involving the determination of which proposed mergers should be permitted and which should be blocked. In December 2023, the US Department of Justice and Federal Trade Commission issued new merger guidelines that revised their 2010 Horizontal Merger Guidelines and extended coverage to vertical and other types of acquisitions that had previously received modest attention.

This article examines how economic analysis in merger guidelines can be improved. For concreteness, it focuses on the 2023 Merger Guidelines and their differences from the 2010 Horizontal Merger Guidelines regarding horizontal

■ *Louis Kaplow is the Finn M. W. Caspersen and Household International Professor of Law and Economics, Harvard University, and a Research Associate at the National Bureau of Economic Research, both in Cambridge, Massachusetts. His email address is meskridge@law.harvard.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20241413.

mergers (that is, mergers between competitors), the primary target of merger enforcement for decades. Although US merger guidelines set forth only nonbinding protocols, they influence merger proposals by describing the regulatory treatment that firms can expect and affect the conduct and outcome of merger litigation by shaping courts' analysis.

Fundamentally, merger guidelines are an application of decision analysis to the determination of whether the expected dangers of proposed mergers exceed prospective gains. Accordingly, we would expect them to devote considerable attention to both the costs and the benefits of proposed mergers. Yet merger guidelines focus almost exclusively on potential harms, giving far less attention to possible benefits, an imbalance magnified by the 2023 Merger Guidelines. Indeed, this is so under the new Guidelines even with respect to non-horizontal acquisitions, which are vastly more common and had long been seen as more likely to be beneficial (as Shapiro discusses in this issue). Closely related, merger guidelines deviate from standard decision analysis regarding information processing and decision-making.

The 2023 Merger Guidelines focus almost entirely on anticompetitive effects. Industrial organization economists have developed models and empirical techniques for predicting the extent to which horizontal mergers are likely to raise prices, or otherwise worsen offers to counterparties, by reducing head-to-head competition between the merging firms. These methods have been increasingly incorporated, reaching a peak in the 2010 Horizontal Merger Guidelines, whereas the 2023 Merger Guidelines pause and, in some respects, reverse course (as Francis discusses in this issue). In any event, all modern merger guidelines deviate significantly from best practices in their focus on market shares that depend on an incoherent market definition apparatus.

Deviations between modern merger guidelines and proper economic analysis are also evident in the treatment of potential efficiencies generated by different types of acquisitions. Largely omitted—and also receiving little attention in industrial organization economics due to increased specialization—are important branches of economic inquiry, notably, the theory of the firm, contract theory, and organizational economics (fields associated with several Nobel prizes) as well as more applied work often conducted at business schools. Also absent are fundamental teachings of economic theory regarding entry and exit and, more broadly, how investment, acquisitions, and other dynamic behavior move resources from lower to higher-value uses in a well-functioning economy.

In the United States, economic thinking has been central to antitrust analysis for over a century, including that undertaken by courts. Indeed, antitrust is a realm in which economic policy analysis had an earlier and greater influence than in other domains. Yet the role played by economics in merger guidelines has been insufficient and may now be fading, even as pertinent economic knowledge continues to grow. Economists can do more on this front to contribute to policy and practice, as further developed in Kaplow (2024b). Product improvement usually boosts demand, even by governments.

Decision Analysis

The decision whether to prohibit or allow a merger—based on the magnitude of any predicted diminution in competitive forces relative to predicted efficiency gains—is best understood in a decision analytic framework. Surprisingly, modern merger protocols deviate from this familiar methodology in important ways. The 2023 Merger Guidelines widen the gap by reframing the antitrust agencies' central question as one of identifying any possible harm rather than determining which mergers are likely to be harmful rather than beneficial overall.

Determination of whether potential costs exceed expected benefits in principle devotes symmetric attention to both sides of the balance. Yet the 2023 Merger Guidelines devote just over 3 pages out of 50 to factors that may favor permitting a merger; the ratio in the 2010 Guidelines, although twice as high, was still only 5 pages out of 37. The structure of the analysis also stands out: Anticompetitive effects are considered first and exhaustively. The merger will be presumptively challenged if these effects exceed some unstated threshold, but otherwise it will be permitted without paying any attention to the merger's benefits, including the possibility that they are nil. When some anticompetitive danger is identified, however, one then proceeds to consider efficiencies (and the potential for postmerger entry) for the first time, and the merger will be blocked unless those outweigh or otherwise nullify the anticompetitive effects.

Consider an analogous process in the context of drug review or the promulgation of environmental regulations: First, fully analyze costs (side effects or expenses of implementing different technology) and approve the drug or promulgate the regulation if they do not exceed some preestablished threshold, without regard to benefits. Second, if the costs do exceed the threshold, then ban the drug or withdraw the regulation unless the benefits are sufficiently large. Or one might instead reverse this, examining only benefits first (and so forth). Given the vast heterogeneity in possible costs and potential benefits, this would be an exceedingly strange way to proceed.

Yet, under official merger protocols, this is purportedly how merger review is conducted (although one suspects that competition agency economists deviate, perhaps substantially, from the agencies' own pronouncements in this regard). This sequential siloing of costs and benefits also reflects modern US court cases. It should be noted, however, that judicial doctrine has followed modern merger guidelines to some extent and, relatedly, how the government chooses to present merger challenges in court. In discourse about merger review, economists and lawyers largely take this unusual framework for granted.

As a matter of sound analysis, if one wishes to know which way a scale tips, it makes no sense to zoom in tightly on one side of the balance until the analysis of that side is complete. Even a quick peek at the other side may lead to a quicker decision. It may also lead to a better one. Many potentially problematic mergers are never challenged, and some that are challenged lose in court: because of the difficulty of establishing a proposed merger's future effects, the threshold for anticompetitive effects is not found to have been exceeded. But even a quick look at proffered merger efficiencies would sometimes reveal them to be negligible or pretextual, in which case the appropriate threshold to challenge the merger should be lower. The relevant threshold for a drug's side effects depends greatly on what it aims to treat—skin rashes or life-threatening disease—and its plausible efficacy in doing so. Likewise, a million-dollar compliance cost per firm for a pollution regulation is large when the regulation only modestly reduces a pollutant that is a moderate irritant, but the same cost may be trivial if significant discharges of a serious toxin would be averted.

There are further problems of basic logic with merger guidelines' sequentially siloed approach. Consider two hypotheses: that the merger is proposed because it will enable the merged firm to raise prices, and that it is instead designed to achieve efficiencies, such as economies of scale. For information to help distinguish between these two hypotheses, it must bear differentially on them, notably, by being more likely under one hypothesis than under the other. However, if one confines analysis to a single hypothesis, it cannot be determined which information raises rather than lowers its relative probability of being true.

Additional considerations undermine this protocol. A central basis for inference uses the merging firms' rationality constraint: if the profitability of a merger arises from some combination of anticompetitive effects (higher prices) and efficiencies (lower costs), then any information about either type of effect revises the appropriate inference about the other. This interdependence reinforces the foregoing logical point about differential likelihoods. Moreover, much information pertains directly to both hypotheses. For example, information on firms' cost structures is relevant to demand estimation, competitive interactions, and efficiencies. And even if some information bears primarily on one or another hypothesis, it is almost never optimal to exhaustively examine only one before turning to evidence on the other. Sequential acquisition of information should be guided by the diagnosticity-to-cost ratio: prioritize information that, for a given cost, is most diagnostic and, for a given diagnosticity, is the cheapest. In addition, at every step, through the final decision, one should use all available information as it relates to any relevant issue to triangulate in reaching tentative or ultimate decisions about the expected effects of a proposed merger.

To a remarkable degree, these simple insights from decision analysis are violated in merger guidelines and implicitly contradicted in many policy analyses of how merger review should be conducted.[1] The sequentially siloed protocol may make some sense in early screening because, if it is fairly obvious that a proposed merger does not raise any anticompetitive concerns, the merger can be cleared promptly. But it makes no sense beyond that. As already suggested, the protocol is problematic with horizontal mergers even if one believes (as some do) that

[1] Most merger analysis adopts a decision-analytic frame, whereas a mechanism design approach (Kaplow 2011) is appropriate, for example, when assessing the impact of anticipated merger stringency on prior investment by potential targets (considered below, in the discussion of entry).

efficiencies are rarely substantial in this setting—because examining them up front would then justify, when that prior is confirmed, a lower challenge threshold regarding anticompetitive effects. Moreover, because the 2023 Merger Guidelines extend this protocol to all manner of acquisitions and not just horizontal ones, downplaying efficiencies not only infects a broader domain but also covers settings in which efficiencies are more likely to be dominant.

Anticompetitive Effects

Throughout the developed world, merger protocols advance two parallel methods for predicting whether and how much horizontal mergers will raise prices (or otherwise harm counterparties): (1) a direct approach that uses the most pertinent economic models, analysis, and evidence, and (2) a market definition approach that "defines" a so-called relevant market from which anticompetitive effects are then to be presumed in light of the merged firm's and other firms' market shares therein. The 2023 Merger Guidelines contain important changes in both: on one hand, they elevate the importance of market definition by strengthening and augmenting presumptions from market shares; yet, on the other hand, they seemingly advance a preference for abandoning conventional market definition altogether, privileging direct analysis. Both changes are best understood after reviewing each of the two methods, beginning with the direct approach (Kaplow 2024b).

Direct Approach

Industrial organization economists—and most merger guidelines, beginning in 1992 in the United States—distinguish between unilateral and coordinated effects of horizontal mergers. Unilateral effects arise because the merged firm will recapture some of the lost profits from price increases on either firm's products. If firm *A* raises its price in a setting with differentiated products, it will lose sales to firms *B*, *C*, *D*, and so forth, all of whom have equilibrium prices above marginal cost. But when firm *A* acquires firm *B*, it will then itself benefit from (that is, recapture) some of those lost sales, and thus firm *A* has an incentive to increase the price of its product by more than it did premerger (and similarly for the price of firm *B*'s product). As explained by Farrell and Shapiro (2010), the magnitude of this incentive is determined by the "diversion ratio" from *A* to *B* (the portion of lost sales from raising the price of *A* that go to *B* rather than to *C*, *D*, and so forth) and the relative price-cost margins on *A* and *B* (the higher the margin on *B* relative to that on *A*, the greater the profit recapture). When a merger is under review, these magnitudes may be estimated by drawing on information in the merging firms' internal documents, interviews with industry players (such as large buyers), and econometric analysis of data (for example, Berry, Levinsohn, and Pakes 1995), including through the use of merger simulations (Baker and Bresnahan 1985; Nevo 2001).

The assessment of coordinated effects—firms' ability to recognize the effects of their actions on each other and hence orchestrate higher prices—has a longer history but poses greater analytical and empirical challenges. Stigler (1964) informally examined factors bearing on the ability of firms to coordinate in the first place as well as to detect and punish cheating. Friedman (1971) launched the use of supergames (repeated one-shot games) to model coordination, and Green and Porter (1984) combined this method with some of Stigler's informal insights. However, identifying coordination, whether achieved through secret meetings or more subtle means, has proved difficult, as has determining the relative contributions of different factors (Marshall and Marx 2012; Kaplow 2013; Harrington 2017). Predicting how particular mergers will influence the ability and extent of firms' coordination is even more elusive, although there is some consensus on the relevant considerations, many of which are noted in modern merger guidelines.

Market Definition Paradigm

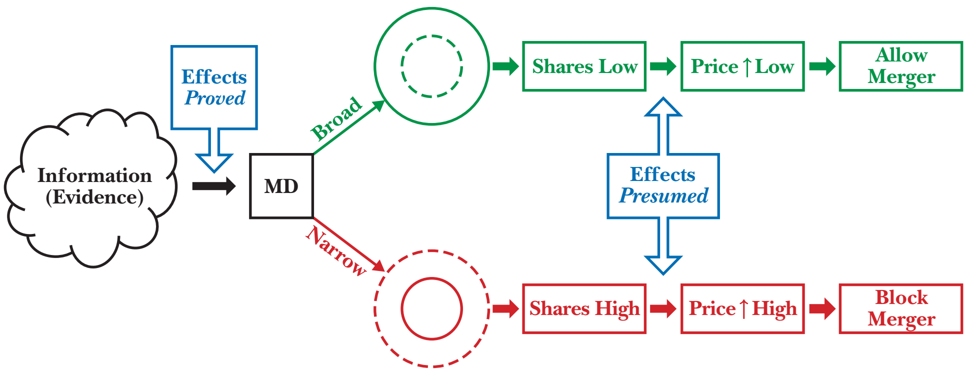
Consider now the alternative approach to merger assessment: the market definition paradigm, an invention of antitrust practice that does not really exist in industrial organization economics (Fisher 1987; Kaplow and Shapiro 2007). This method defines a metaphorical "market," computes market shares in that market, and presumptively blocks (or allows) a merger if various market share measures exceed (or fall below) posited thresholds. The 2023 Merger Guidelines presumptively challenge a merger with a postmerger Herfindahl-Hirshman Index (HHI) above 1800 and a ΔHHI (increase in the HHI) above 100. (The HHI is the sum of the squares of the market shares—in percentage points—of the firms in the "market.") Other modern merger guidelines are broadly similar, although some have weaker presumptions.

For example, if there are five canoe manufacturers each with a market share of 20 percent in the "canoe market," the HHI is 2000 (5×20^2). If two were to merge, the naively computed postmerger HHI is 2800 ($(3 \times 20^2) + 40^2$), implying a ΔHHI of 800. This merger would presumptively be blocked under most modern merger guidelines. However, the merging parties may object to this Narrow market definition, contending that the relevant market is a Broad one including all small, unpowered pleasure craft (rowboats, kayaks, and more). In that market, where canoes are only a modest slice, the postmerger HHI may be more like 175 and the ΔHHI merely 50, so the merger would be allowed to go forward. Hence, battles over market definition can be decisive.

Consider the depiction of the market definition paradigm shown in Figure 1. Start at the right of the figure. It is typically said that anticompetitive effects may be *presumed* from the market shares in the chosen market. Yet this ignores that the market had to be defined in the first place (the decision node denoted "MD" for market definition). If we suppose that markets are chosen functionally, then the information and analysis feeding into the market definition decision should consist of the best evidence and methods for predicting anticompetitive effects. That is, the effects must first be *proved* in order to define the market in the first place. This protocol is patently circular.

Unfortunately, the market definition paradigm is far worse, which can be appreciated by making the decision rule explicit. Let our criterion be to choose

Figure 1
Market Definition Paradigm



Narrow or Broad so as to minimize the resulting error: that is, choose Narrow if and only if the error we make when we choose Narrow is smaller than the error we make when we choose Broad. But neither error can be determined in the first place except by reference to some best estimate of the merger's effect. This estimate, in turn, is precisely the result of the first method: using economic analysis to predict the merger's effect as best we can (without regard to any market definition). Clearly, this estimate must be in hand *before* we define the market, confirming the circularity claim.

We can also see that this method introduces further, avoidable prediction error. How great is this supplemental error? The answer is conceptually indeterminate because both the error when we choose Narrow and the error when we choose Broad are generally unspecified. The latter quantity is especially strange because there do not exist models of "broad" markets in which market shares play a meaningful role. By contrast, more can be said about "narrow" markets. The results are remarkably discouraging for the market definition approach to the assessment of horizontal mergers' anticompetitive effects.

For concreteness, consider the Cournot model of quantity competition with homogeneous goods as it has been applied to the analysis of horizontal mergers (Farrell and Shapiro 1990). Begin with a naïve version that employs several simplifying assumptions, including (incorrectly) that firms' market shares do not change as a result of the merger—a standard premise in merger guidelines' computations of HHIs and ΔHHIs. Then, the industry-wide, average, output-weighted markup is HHI/$|\varepsilon|$, where ε is the elasticity of demand for the homogeneous goods. Thus, the merger raises price by ΔHHI/$|\varepsilon|$, which is obtained by subtracting the premerger value of the formula from its postmerger value.

With this background in mind, consider two different mergers. The first merger is as follows: (1) it just passes the hypothetical monopolist test that is invoked in modern merger guidelines (meaning that a hypothetical monopolist of the posited market would raise price just over 5 percent); (2) it has a postmerger HHI of 1801 in that homogeneous goods market; and (3) it has a ΔHHI of 101. Features (2) and (3) just trigger the 2023 Merger Guidelines' challenge presumption, even though it can be shown that this merger would raise price trivially, approximately 0.06 percent (that is, 6 hundredths of a percentage point).[2]

The second merger is one to a full monopoly involving different homogeneous goods, and let us suppose that the first step of the hypothetical monopolist barely fails. Straight from the hypothetical monopolist test itself (using the same trigger of 5 percent), we know that our actual merger to monopoly would raise price by approximately 4.9 percent. Yet this merger may well not be challenged under merger guidelines' machinery because the hypothetical monopolist test in this case requires broadening the market, and in that market (let us suppose) the merged firm's market share and overall market concentration plummet, falling far below standard thresholds for a challenge (like in the above example with canoes, when other types of watercraft were included). Combining these two cases, the 2023 Merger Guidelines' apparatus would not trigger a challenge to a merger to monopoly that raises price by *more than 80 times as much* as another merger that is presumptively illegal ($4.9/0.06 \approx 82$).

Nocke and Whinston (2022) employ more sophisticated analysis that takes account of the endogeneity of market shares.[3] In their model, the premerger, equilibrium (endogenous) market shares provide information on the firms' underlying marginal costs, and those (combined with other informational inputs) enable a prediction of the postmerger equilibrium. Employing the pertinent simplifying assumptions, which include premerger symmetry among firms, the predicted price increase from a merger is a function of the square root of ΔHHI and the elasticity of demand. Note first that, even limiting attention to homogeneous goods markets that would be validated by the hypothetical monopolist test, the magnitude of this elasticity can exceed 15;[4] hence, the elasticity can differ by more than an order of magnitude (even if considering only elasticities above one). Now, because

[2] For a back-of-the-envelope indication, a hypothetical monopolist in a completely unconcentrated homogeneous goods market would raise price 5 percent if the magnitude of the elasticity were 20, so a ΔHHI of 100 (which should be 0.01 for this formula, which does not use percentage points like the colloquial HHI) would raise price by $0.01/20 = 0.0005$. Because the example in the text has a baseline HHI of 1700, the corresponding elasticity is somewhat lower and the price increase slightly higher.

[3] Some readers may wonder about the market definition paradigm because the prediction of price effects from market shares is reminiscent of the structure-conduct-performance paradigm (which regressed prices or profits on market shares). This approach has long been in disrepute due to the endogeneity of market shares (Demsetz 1973). Careful modern analyses, like Nocke and Whinston (2022), work with model primitives that are exogenous, solve for pre- and postmerger equilibrium prices, determine the difference, and, as a last step (which is possible analytically only for special cases involving cost functions and demand), substitute the (endogenous) premerger market shares into the formula.

[4] See the calculations in note 2.

the ΔHHI enters the formula as its square root, the critical value of the challenge threshold for the ΔHHI *varies by more than two orders of magnitude.*

These simple Cournot illustrations are sufficient to indicate that the market definition paradigm distorts and discards information to such an extent that enormous prediction errors can arise. As explained above, these errors are not those inevitable in the prediction of mergers' effects but are supplemental: they are deviations relative to the best predictions we would make using the same information set and the best tools available.

Market definition fares no better in predicting unilateral effects with differentiated products. The core method (described above), which focuses on diversion ratios between the merging firms' products and the price-cost margins on each, is understood to operate without regard to market definition. (Indeed, Farrell and Shapiro's 2010 title refers to the method as "An Economic Alternative to Market Definition.") Nor is market definition useful in analyzing coordinated effects: regardless of the hypothetical monopolist test or any other protocol, the homogeneous goods market is usually the proper focus because that is where price coordination is most likely to be feasible. And even in this market, HHIs and ΔHHIs are of little use.

Merger Guidelines

Return now to the changes contained in the 2023 Merger Guidelines. The first is that market share presumptions are made more powerful in generating merger challenges. Guideline 1 lowers the joint challenge threshold from a postmerger HHI/ΔHHI of 2500/200 in the 2010 Guidelines to 1800/100 (where they had been before the 2010 revision). It also introduces a new, independent basis for challenge: when the merging firm's postmerger share exceeds 30 percent and the ΔHHI exceeds 100—blocking, for example, a 28 percent / 2 percent merger. Obviously, both of Guideline 1's presumptions require market definition, for the posited market shares are deemed to be those in the "relevant" market.

The second, opposing, and potentially more sweeping change appears toward the end of the 2023 Merger Guidelines, regarding the market definition process itself. There, use of the hypothetical monopolist test—the central means of market definition in all US merger guidelines since 1982 and followed in much of the world—is demoted from the primary method to last place on a list of four. It is preceded, in third place, by the use of "practical indicia" from the 1962 *Brown Shoe* Supreme Court decision—in rough terms a set of intuitive factors (such as public recognition, product characteristics and uses, and distinct customers) indicative of the degree of substitution.[5]

[5] In the courts and other informal antitrust discourse, it is widely believed that markets should be broadened to include close substitutes (indeed, beginning in early cases, cross-elasticities are mentioned explicitly). This inclination fails to appreciate that some version of the market elasticity is most appropriate to use. After all, the market elasticity of demand is a revenue-share weighted sum of the cross-elasticities with respect to all other goods in the economy, so it is a vastly superior indicator of total substitution than is picking a handful of goods with "high" cross-elasticities, and without regard to the proper revenue weights, while ignoring everything else.

The top two positions are occupied by "direct evidence." First place is held by direct evidence of substantial competition between the merging parties, which sounds like how economists analyze unilateral effects without regard to market definition. (Material on just that appears in the immediately preceding section of the 2023 Merger Guidelines.) Yet direct evidence is not a *means of defining* a market but instead a *replacement* for market definition.

The 2010 Horizontal Merger Guidelines were seen as part of an ongoing process of shaping merger analysis in an ever more explicit economic mold, particularly regarding unilateral effects (as Francis discusses in this issue). However, they stopped short of completely embracing correct economic methodology—as reflected in their continued endorsement of the hypothetical monopolist test and their use of market share threshold tests. Many followers of the US antitrust agencies suppose that, in their internal analysis, economists in fact place significant weight on economic methods and then reverse-engineer market definitions to some degree. But that procedure was never the official position, and it is not how the agencies and merging parties typically litigate in court. The 2023 Merger Guidelines may signal a radical shift. Moving the use of direct economic analysis to the top of the list may encourage merging parties to alter their presentations to agencies and courts. Because courts have largely followed the agencies' lead by adopting the methodology of past guidelines, this revision might ultimately produce a significant change in merger adjudication, which could in turn spill over to other areas of antitrust that rely on market definition. This is more a possibility than a prediction, from the perspective of perhaps the most ardent critic of the market definition paradigm (Kaplow 2024b).

In closing, it is worth reflecting on Lerner's (1934) classic article. Although known today for the Lerner index of market power, the article's primary aim was to nip in the bud what he regarded as the contemptuous practice of using market definition and market shares to assess market power. He used scare quotes to emphasize the vacuity of attempting to define "commodity" and "industry." Whether Lerner's hope will be realized by the centennial of his article remains to be seen.

Efficiencies

Merger Guidelines

Acquisitions of assets—the actual domain of US "merger" law, Clayton Act Section 7—are a principal means by which resources are reallocated from lower- to higher-value uses in a well-functioning economy. Hence, a broad prohibition on asset acquisitions, even just significant ones, would obstruct or eliminate many of these flows and prevent the creation of sophisticated firms as we know them. It is not surprising, therefore, that modern merger guidelines typically limit their scope to particular types of mergers and also contain some sort of efficiencies defense that contemplates justifications for what may otherwise be regarded as anticompetitive acquisitions, or at least reasons that some acquisitions should not be deemed anticompetitive in the first place.

From a distance, both the 2010 and the 2023 Merger Guidelines embody this perspective in their seemingly similar sections on efficiencies. But a closer look at the modifications in the 2023 Guidelines suggests a major shift. First, the 2023 Guidelines drop longstanding declarations that many mergers promote efficiency, which explicitly acknowledged the downside to prohibiting some mergers.

Second, unlike the 2010 Horizontal Merger Guidelines, the 2023 Merger Guidelines cover not only horizontal mergers but also vertical and even some unrelated acquisitions of assets.[6] The 2010 Guidelines already embodied a highly skeptical stance toward efficiencies, which many thought to be warranted when (1) only substantial horizontal mergers were covered, (2) a high challenge threshold was employed (that implicitly credited substantial efficiencies), and (3) large horizontal combinations were not thought to generate significant merger-specific efficiencies very often. But this rationale does not apply to the full gamut of asset acquisitions that are now included—with a limited, one-size-fits-all efficiencies provision that was copied (but narrowed) from the prior provision developed only for horizontal mergers.

The 2023 Merger Guidelines also contain a more specific, related change: they exclude wholesale any efficiencies that are "outside" the market that is the locus of anticompetitive effects.[7] But when resources are redeployed from a lower-value use to a higher-value one, competition often will fall in the former domain and rise in the latter. These outcomes can arise when a firm acquires a plant to convert the facility to a new type of operation, acquires land to remove existing buildings to make way for a chip fabrication plant, or acquires a professional services firm to redeploy its employees' talents to clients in a different sector—and in countless other ways. If merger assessments count only the loss (less competition due to the extinguished activity) and never the gain (from redeployment to the new activity), then this change in the 2023 Merger Guidelines—combined with their extension to a much broader array of acquisitions—could have an enormous effect (as Shapiro discusses in this issue).

Economic Analysis

What does economics teach us about the role of merger efficiencies? Begin with the central point that resource allocation in a market economy ubiquitously involves acquisitions (which is to say, purchases) of all manner of assets—and sometimes of entire firms, usually quite small but sometimes very large. Next, efficiencies are one side of the balance of harms and benefits of mergers, and they are also

[6] This article does not address the short-lived 2020 Vertical Merger Guidelines, which were much more limited in scope than are the vertical sections of the 2023 Merger Guidelines.

[7] Although the text of the 2010 Horizontal Merger Guidelines can be interpreted similarly, the relevant passage contains a key footnote (p. 30, n. 14), removed in the 2023 Merger Guidelines, that reads as follows: "In some cases, however, the Agencies in their prosecutorial discretion will consider efficiencies not strictly in the relevant market, but so inextricably linked with it that a partial divestiture or other remedy could not feasibly eliminate the anticompetitive effect in the relevant market without sacrificing the efficiencies in the other market(s)."

central to inferences about anticompetitive effects (as explained in the discussion of decision analysis). Indeed, one may ask: If mergers are typically assumed to be proposed because of some combination of efficiencies and anticompetitive effects, and if efficiencies are to be ignored entirely in a wide range of settings, what is there left to analyze? The very attempt to undertake an acquisition would then imply that it was anticompetitive, and any balancing exercise would be easy because there would be no admissible costs to blocking broad swathes of mergers. The changes in the 2023 Merger Guidelines can aggressively be interpreted in just this way, although it remains to be seen whether the agencies will attempt to do so and whether the courts would be at all receptive.

To pursue the economic analysis of efficiencies more directly, let us begin with the standard requirement that is referred to as "merger specificity": if proffered efficiencies are to *justify* a merger that might otherwise be detrimental, then the merger should be *necessary* to generate the efficiencies. Often, as merger guidelines suggest, the contemplated alternative is contractual. For example, if the target firm would benefit from some skill that the acquirer will extend to its operations, we should ask whether the target can instead obtain the requisite boost by contracting with a third party that is not a competitor (efficiency in payroll processing can be achieved by outsourcing the task). If there is a significant scale economy associated with new high-tech medical scanning equipment, we can consider whether two merging hospitals might instead form a joint venture with respect to just that asset, remaining as independent competitors for all other services.

This core idea is sound, but it can be difficult to apply beyond such simple examples. Broadly, the merger-specificity question is a species of the inquiry central to economic analysis of the theory of the firm (Coase 1937; Williamson 1975, 1985; Grossman and Hart 1986; Hart and Moore 1990; Simon 1991; Holmstrom and Milgrom 1994): When, why, and how is it that bringing some activities "inside" a "firm" can generate benefits that cannot be obtained "outside" the firm via "contracts"? Jensen and Meckling (1976, p. 311) famously stated that a firm is a "legal fiction" that is nothing more than a "nexus for contracting relationships," and Alchian and Demsetz (1972, p. 777) referred to the belief in commonly advanced differences between firms and market relationships as a "delusion."

The answers have to be found by engaging with the substance rather than invoking labels.[8] Hence, we need to analyze hold-up problems, asymmetric information, coordination of complementary activities, and other phenomena central to the institutional organization of the economy.

As a consequence of increasing specialization, both the aforementioned theoretical developments (associated with several Nobel prizes) and more applied work

[8] Often overlooked is the fact that firms are often organized internally in a decentralized manner that mimics markets, and that contracts (for example, involving franchises or supply chains) often entail substantial control assigned to one of the parties (Stinchcombe 1990; Bernstein 2015). Moreover, it is relevant whether contractual alternatives that may well involve complex interactions between competitors will themselves be regulated (which, in practice, they usually will not be) so as to prevent anticompetitive effects similar to those from a horizontal merger.

(often conducted at business schools; for example, Chandler 1990) have largely been missing from the economic analysis of antitrust policy, including of mergers. The fact that only about 6 percent of the 2010 Guidelines and just over 2 percent of the 2023 Guidelines address efficiencies even though they (crudely) might be thought to constitute half of any plausible assessment is an indication of this underdevelopment. Essentially none of the core ideas from the theory of the firm, contract theory, and organizational economics (March and Simon 1958; Gibbons and Roberts 2013) appear in modern merger guidelines or most commentary on the subject. Likewise for modern economic research on productivity (De Loecker and Syverson 2021).

To illustrate some of these issues, consider two settings (discussed in more detail in Kaplow 2024b). First, take the familiar case of economies of scale as a purported justification for some horizontal mergers. Suppose that an industry initially has ten symmetric firms, each consisting only of a single plant of minimum efficient scale. Then, an innovation doubles minimum efficient scale. The firms propose to pair off and merge down to five firms, offering an efficiencies defense based on these heightened scale economies. A near-consensus view is that this justification is weak, for the industry can sort itself out without any mergers. However, if we still end up with five firms (half exit; the other half double in size), society not only enjoys the benefits of the scale economies without allowing the mergers but also suffers the same anticompetitive effects from having only five firms rather than ten. (More subtle analysis, beyond the scope here, suggests benefits to blocking such mergers, particularly when the efficiencies are uncertain to the regulator, because the conditionality of the reshuffling of positions tends to be superior.)

Often overlooked in such analyses is that most firms are not stick figures that consist of single assets, like plants. A one-plant firm making a single product needs supply chains, financing, employees (requiring hiring, training, and benefits), distribution channels, marketing, customer relationships, and perhaps a reputation for quality and fair dealing. Moreover, a single retail store may offer thousands of products, and a single plant may manufacture some of its intermediate inputs and use them to produce different, related products. And many important firms are vastly more complex. Even in simple cases, firms are often bundles of complementary assets and activities, including contracts and other relationships, that together can involve significant going concern value: the whole is worth more than the sum of the parts.

Recognizing this basic feature of most significant firms, return to our contemplated prohibition of the pairing-off mergers, leaving half the firms to double in size and the other half to exit. How do the former firms expand their complementary assets and capabilities? And what happens to the going concern value of the respective assets and relationships of the latter firms? If the surviving firms can acquire those otherwise stranded sources of value from the exiting firms, those are, well, acquisitions, which we are contemplating to be prohibited. Hence, it may be that, if and when there are efficiencies from horizontal mergers, those efficiencies—themselves involving complementary assets preexisting within the merging firms—are specific to the merger after all. Interestingly, it is the anticompetitive effects that may not be merger specific, for if the pertinent acquisitions are prohibited, we still end up with only five firms rather than ten, but with forgone efficiencies and thus higher costs because the exiting firms' value is lost while the remaining firms must spend afresh to replace that lost value.

For the second setting, consider another, much broader class of situations that is particularly relevant under the 2023 Merger Guidelines due to their coverage of vertical and even some unrelated acquisitions of assets. Complementarities are widespread in the economy and often make it efficient to combine different types of assets in single firms. Contemplate the opposite world: Every narrowly defined asset and activity is required to be entirely independent. An automobile would have no two components produced by the same firm; every employee would be an independent contractor hired separately for each task; and all supplies, assembly, and sales would be through separate, spot market transactions. Perhaps business software, including operating systems, would have each line of code or code module independently owned; assemblage for each user would be accomplished through myriad spot market contracts. Imagine the hold-up problems, the information asymmetries associated with contracting, the difficulties in coordination of the co-evolution of hundreds or thousands of interrelated features, and the problem of Cournot complements (and associated "double" marginalization) when there are hundreds or thousands of pieces rather than just two. Not to mention even more mundane transaction costs, such as finding and contracting with each supplier and separately transferring funds to dozens or thousands of counterparties for a single, simple action.

Complex integration of complementary activities is often essential. Broad prohibition would entail massive value destruction. But how do acquisitions fit in? First, as noted, "merger" regulation actually covers all acquisitions of assets. Most internal growth involves countless asset purchases. Even that which is produced internally requires buying inputs—and often large purchases, such as when a chip fabrication plant or charging station network is constructed.

Second, and closer to any serious boundary of contention, firms often find it more efficient to acquire clusters of assets (some of which may exist as other firms) rather than to assemble everything from scratch. If every firm developing a new product or entering a new line of business was required to halt all such activity until each obstacle could be overcome through purely internal means that are limited to hiring each employee and creating every little thing separately—when it might instead be possible to acquire the capability from an existing entity—much innovation and even fairly ordinary, gradual adaptation would be stymied. Many firms have distinctive skills that other firms lack and cannot easily duplicate independently. If the contrary were true, Walmart's information technology, distribution, and more could readily be imitated, and likewise for countless other highly successful firms in many sectors of the economy. Hence, as in the foregoing example with simple scale economies joined with many complementary assets and relationships, potential economies of scope that involve complementary capabilities may sometimes be important even in what are usually regarded as purely horizontal mergers. And they are typically central to vertical and other types of cross-market acquisitions.

This discussion is only meant to be suggestive. Economic research and antitrust practice would benefit from renewed efforts at cross-fertilization if we wish to enhance merger review in the short run and improve merger protocols in the future. Many of the central questions and decades of associated research are largely off the radar screen in antitrust literature, government guidelines, and merger review. Nor have those in related disciplines undertaken the work necessary to apply their findings to merger analysis or most other areas of antitrust.

There is another important hurdle that both the 2010 and the 2023 Merger Guidelines rightly emphasize: verifiability. It is difficult for a reviewing agency or court to figure out whether merging parties' efficiency claims are compelling or pretextual. Although potentially significant efficiencies may lie in many places, it is another matter to assess their importance in the review of particular mergers. Literature on the theory of the firm, contract theory, organizational economics, and productivity elaborates subtle ideas that have little connection to modern training in industrial organization economics or to that of antitrust lawyers. In the United Kingdom, it is more common to include someone with industry-specific business expertise from the outset of important investigations. By contrast, the United States and many other jurisdictions' competition agencies do not staff merger review in this manner. They contact some industry players, but occasional, informal discussions with volunteers (who may have ulterior motives) are woefully insufficient. Particularly for the analysis of efficiencies—but also for examining anticompetitive effects and entry—agencies really need to have on-staff or substantial, early access to on-call expertise if they are to collect the right information and analyze it effectively. Consider how disturbed we would be to see an environmental regulator attempting to undertake cost-benefit analysis with a staff of public health experts and lawyers, but no engineers to illuminate industry submissions regarding the costs and efficacy of proposed technological solutions.

Entry

Entry and exit are central to the functioning of a market economy, one of the basic means by which resources are channeled to their highest value uses. In particular, entry is one of an economy's ways of restoring or otherwise augmenting competition and of launching new activities that contribute greatly to social welfare.

In analyzing mergers, it is helpful to distinguish entry at two different points on a timeline, where a proposed merger is placed at time one. At time two, we may have entry induced by a merger that would raise prices, which makes entry more profitable than it had been premerger. At time zero, we may have entry induced by the prospect of a future buyout; such entrants, that is, may hope to be targets of future acquisitions at premium prices. The 2010 and 2023 Merger Guidelines treat the former similarly and the latter perhaps quite differently, both largely in ways that do not reflect proper economic analysis.

Postmerger Entry

Regarding entry at time two, modern merger guidelines focus on whether it is likely to be sufficient to largely eliminate any anticompetitive price increases that the proposed merger would otherwise cause. However, short of fairly extreme cases (some of which do exist), this will not occur because greater postmerger profitability, caused by postmerger equilibrium prices that remain elevated, is needed to induce entry that was unprofitable beforehand (Werden and Froeb 1998; Kaplow 2023), a conclusion that simulations bear out (Caradonna, Miller, and Sheu forthcoming). Hence, the skeptical, stringent stance of the 2010 and 2023 Merger Guidelines in this regard is appropriate.

But asking whether postmerger entry will almost completely offset any price increase is usually the wrong question. Instead, we should consider how the prospect of entry that partially mitigates price increases may render an otherwise anticompetitive merger unprofitable, keeping in mind that the merged firms' quantity reductions generate positive externalities on other firms, including such entrants.[9] If firms only propose mergers that they expect to be profitable, and if profits arise either from sustained price increases or from efficiencies, we can see that easier postmerger entry shifts inferences about a proposed merger's motivation away from anticompetitive effects and thus toward efficiencies. (This basic logic, unfortunately, is largely missing throughout modern merger guidelines.) Moreover, if merging firms expect to be sufficiently more efficient that they would lower prices, then ease of postmerger entry is irrelevant and the sometimes associated ease of exit may render such a merger more profitable than otherwise, strengthening the inference that the proposed merger is motivated by efficiencies. Note how this consideration reinforces the above suggestion that merger review should triangulate—using evidence bearing on anticompetitive effects, efficiency, and entry—rather than consider each subject in a sequentially siloed fashion.

Merger guidelines' analysis of postmerger entry also ignores the direct effect of such entry on social welfare (Spence 1976; Dixit and Stiglitz 1977; Mankiw and Whinston 1986). First, entry tends to be socially excessive in homogeneous goods industries because the fixed costs of entry are covered by positive margins on sales taken from other firms that sell at prices in excess of marginal cost (the "business-stealing externality"); induced entry burns more resources in covering fixed costs than any gain from enhanced competition. It is even possible that the main social cost of a merger is its inducement of additional, inefficient entry (Kaplow 2023). Second, entry can be socially insufficient when entrants' products

[9] Indeed, even without entry, horizontal mergers that generate no efficiencies can be unprofitable because the merging firms may lose more profits from their quantity reductions than they gain from higher margins on what they continue to sell (Stigler 1950; Willig 1991; Werden and Froeb 1998; Whinston 2007).

are differentiated and thus add to variety (or entrants otherwise generate positive spillovers, such as through incompletely appropriated innovation). In this case, some mergers raise social welfare, and even consumer welfare, on account of higher prices' inducement of additional entry.

Entry in Anticipation of a Buyout

Time zero entry, which is induced by the prospect of a subsequent buyout, may be even more consequential, yet its consideration is almost entirely omitted from modern merger guidelines. In many sectors—with pharmaceuticals and digital technology receiving the most attention—there is a steady stream of new entrants, many of whom may be induced to enter in significant part because of the prospect of a subsequent buyout at a premium price. In this setting, prior entrants' existence and capabilities at the time of a proposed merger should be viewed as endogenous to the expected stringency of the merger regime, a core point that is widely ignored. Mergers are part of the dynamic ecosystem that includes entry, investment, competitive interaction, and exit (Ericson and Pakes 1995).

Entry for buyout was analyzed by Rasmusen (1988) in a simple model of a homogenous goods industry, an incumbent monopolist, and other assumptions that together imply that the prospect of subsequent buyout can motivate inefficient entry (fixed costs are consumed but little subsequent competitive benefit ever arises). A strict merger regime, in anticipation, would discourage this excessive entry, indeed, to the benefit of the monopolist who would like to be able to commit to not buying out entrants. This analysis is extended by Gowrisankaran (1999) and Mermelstein et al. (2020), who show that the result can arise in a more sophisticated dynamic framework—but under assumptions that limit the number of incumbents (else there is a free-rider problem) and specify particular timing of entry, investment, production, and bargaining. Interestingly, in this case there is a strong argument for prohibiting acquisitions of what might, viewed at the time of proposed mergers, be viewed as disruptive entrants (Cunningham, Ederer, and Ma 2021). However, the central effect of the prohibition is not more actual disruption but instead to (efficiently) discourage such entry in the first place—thus turning a common objection to stringent merger policy into a social benefit.

A central source of these results is the aforementioned point that entry tends to be socially excessive in homogeneous goods industries. But when entrants invest in complementary activities, contribute to variety, or generate other positive spillovers (such as incompletely appropriated innovation), the prospect of buyout at a premium helps counteract the generic social insufficiency of entry and thus generates benefits beyond those realized by the merging parties. Much innovative investment in start-ups or by other small firms has little prospect of ever producing an independently viable product. A new biotech firm may begin development of a drug or device, or a software firm may design complementary capabilities for an operating system, or a firm may create process improvements, all for activities that such firms could never hope to undertake themselves. Sometimes, the optimal policy would be for them to license their innovations to all, but such is often infeasible or severely limited by difficulties in contracting with asymmetric information or the need for focused incentives. (No pharmaceutical licensee would spend greatly to bring a drug to approval if, at that point, other licensees of the same patent could freely compete.)

As a crude first approximation, therefore, an optimal merger regime may well be strict toward acquisitions of firms whose activities are close substitutes for those of acquirers, but generous toward acquisitions of complements. This differential stringency tends to channel investment away from realms where it is already excessive and toward those where it is inadequate (Gilbert and Katz 2022). More complete analysis of this varied and often subtle setting would be useful, along with the development of empirical evidence to guide its application. Although it will typically be challenging if not impossible for a competition agency to assess many of the relevant factors in its review of specific mergers, its overall policy should reflect these time-zero incentives. Indeed, one might view this setting, which centrally affects innovation and the potential for disruptive competition, as the locus of some of the most important effects of a merger regime. The above suggestion that agencies would benefit from greater in-house or on-call industry-specific expertise is particularly important in reviewing such mergers.

Modern merger guidelines contain essentially no discussion of incentives for entry in anticipation of buyout premiums. The 2023 Merger Guidelines do address acquisitions undertaken by firms that might be regarded as potential entrants into the activities conducted by the target, along with other scenarios involving acquisitions of new entrants. The suggested regulation seems quite strict, which could conceivably ban many classes of buyouts, including those of targets engaged in different activities from those of the acquirer. If investors and firms anticipate such strictness, this central channel of dynamism could be severely curtailed. The failure to include any mention of investment incentives for prospective merger targets constitutes a fundamental departure from sound economic analysis. As with other apparent changes in the 2023 Merger Guidelines, it remains to be seen whether and how the antitrust agencies will pursue these new pathways and, if they do, whether courts would be receptive.

Tools and Targets

Modern merger guidelines and other competition policies reflect certain broader perspectives toward economic policy analysis that differ from those in other applications. Most familiar and somewhat controversial is the focus on consumer (more broadly, counterparty) welfare rather than total welfare. This view is typically motivated by distributive concerns: higher prices tend to have a distributive incidence matching consumption, whereas firm profits tend to be distributed in accordance with the ownership of capital. The 2010 Horizontal Merger Guidelines sought to avoid mergers that cause "firms to raise price, reduce output, diminish innovation, or otherwise harm customers." The 2023 Merger Guidelines refer to the gains resulting from businesses that "lower prices, improve wages and working conditions, enhance quality and resiliency, innovate, and expand choice, among many other benefits." Neither set of guidelines mentions producer surplus. The consumer welfare versus total welfare distinction is perhaps most explicit where both sets of guidelines indicate that they credit merger efficiencies only to the extent that they are passed on to consumers.[10]

In other domains, economists have tended to favor designing specific forms of regulation—whether of the environment, drugs, or safety—so as to maximize total welfare within their domains, leaving distribution to the tax and transfer system. A broad justification is specialization: every branch of government could seek to address every aspect of social welfare, including distribution, national defense, and macroeconomic stability, but nothing close to that is done for reasons of expertise, focus, and accountability. Moreover, tasks are best assigned to agents who control the regulatory instruments best suited to hit their designated targets. The tax and transfer system tends to be the tool most able to address broad distributive concerns.

Discussions of the proper welfare objective of competition regulation are largely uninformed by the past half-century of work in public economics that is most associated with Mirrlees's (1971) analysis of optimal income taxation and Atkinson and Stiglitz's (1976) extension to additional instruments (there, differential commodity taxation). This apparatus has been applied to a broad range of policies, showing how it is a Pareto dominant strategy in benchmark settings to confine distributive considerations to the income tax and transfer system (Kaplow 2024a). This formulation has been shown to favor the pursuit of total welfare rather than consumer welfare in competition regulation (Kaplow 2021). Moving from a consumer to a total welfare standard, combined with a distributively offsetting adjustment to the income tax and transfer schedule, enables a strict Pareto improvement; that is, it makes every income group better off.[11]

Another feature of most analysis in modern merger guidelines, which diverges from some other areas of competition regulation, is its focus on the short run. Both the 2010 and the 2023 Merger Guidelines express an interest in entry and efficiencies only when they will be realized speedily. But long-run equilibrium effects are often more important than short-run effects and can be opposite in sign. Indeed, all investments are losers in the very short run regardless of whether and by how much the present discounted value exceeds zero. A short-run perspective would take a

[10] In another respect, the 2023 Guidelines may depart, perhaps significantly, from the 2010 Guidelines—by emphasizing so-called harms to "competition" (which are mentioned frequently, although the meaning is unclear), with a possible de-emphasis on welfare effects of any kind (as Francis discusses in this issue).

[11] Consider a more familiar application: environmental regulation. Because life and health are far more valuable (measured in dollars) to the rich, relative to the incidence of costs (perhaps roughly proportional to consumption), heavily weighting distributive concerns would favor much weaker regulation. An otherwise optimal carbon tax, for example, would be regressive, but rebating the revenue through a distributively offsetting income tax adjustment would eliminate distributive concerns, making a total-welfare-based cost-benefit analysis appropriate for setting the carbon tax.

negative view of increases in innovation. In other realms of competition policy, a short-run focus would ignore many detrimental exclusionary practices (as one example, predatory pricing benefits consumers in the short run at the expense of the long run). Indeed, even regarding unilateral effects of horizontal mergers in a Cournot model of competition through the choice of quantities, in the short run many mergers lack anticompetitive effects (and fail to be profitable in the absence of efficiencies) even when they are anticompetitive and profitable after investment subsequently adjusts (Berry and Pakes 1993). Analysts need to be careful about artificially truncating the time horizon.

As a practical matter, agencies with finite resources making rapid decisions cannot hope to reliably assess the long-run effects of proposed mergers, so the short-run focus is understandable. Nevertheless, protocols should be designed with long-run considerations in mind. For example, the earlier discussion of incentives for entry in anticipation of a buyout premium emphasizes that the most important consequences of the anticipated stringency of merger regulation involve the direction and magnitude of entrants' prior investments. Even though these anticipatory effects often cannot be gauged for a particular merger proposal, classes of proffered mergers can be identified where a more stringent or permissive approach would be superior.

Note further that, in the long run, consumer and total welfare standards tend to have similar implications. In many models of long-run equilibrium with endogenous entry, exit, and investment (Hopenhayn 1992; Ericson and Pakes 1995; Melitz 2003), expected profits are zero, all interim profits are quasi-rents that recover the costs of prior investments, and price equals average cost. All costs are variable in the long run and ultimately are borne by consumers. From this long-run perspective, the objective of merger and other competition regulation is to prohibit rent-seeking that destroys social value while keeping open channels for resource reallocation and innovation that raise welfare.

Finally, different areas of economic policy analysis give differential attention to problems of the second best, where correcting one market failure might exacerbate another. Such considerations have been more important, for example, in some areas of public economics and international trade, but not in industrial organization. Yet Lipsey and Lancaster's (1956, p. 16) famous article on the general theory of the second best featured the problem of "degrees of monopoly" (markups in many sectors), which was apparently so well known that they felt no need "to review the voluminous literature on this controversy" (see also Lerner 1934).

Research in industrial organization economics, including applications in antitrust, often introduces an "outside good" in order to isolate the sector under investigation (Lancaster 1980; Mankiw and Whinston 1986; Berry 1994). When used for welfare analysis rather than demand estimation, however, we should be concerned by the assumption that this good—corresponding to the rest of the economy—is implicitly taken to be available at a price equal to its marginal cost. After all, there are significant and varying markups in many sectors of the economy (for example, Hall 2018; De Loecker, Eeckhout, and Unger 2020; and the discussions in Basu 2019; Syverson 2019; Shapiro and Yurukoglu 2024).

When ameliorating anticompetitive price increases in a targeted sector, more resources will accordingly flow into that sector, reducing deadweight loss there. But those resources thereby flow out of other sectors that may also be distorted, so the overall welfare impact is less clear. This problem is rendered more complex when one accounts for changes in entry and exit (which, as discussed above, are generally distorted as well) in both the targeted sector and in other sectors, as well as general equilibrium price adjustments that are welfare-relevant in a distorted economy. Welfare analysis of this broader problem has only recently begun (Kaplow 2023) and constitutes another important area where further research can provide important direction, particularly prioritization, for competition regulation, including of mergers.

Conclusion

Merger regulation, as embodied in modern merger guidelines, is a mixed success regarding the use of economic analysis. Compared to the 1960s, economic analysis now has a far greater impact—in the United States and throughout the world—partly reflecting ongoing developments in industrial organization economics. Unfortunately, however, many advances in the field as well as key teachings in other areas of economics have been largely ignored. Some of these shortcomings reflect increasing specialization. Those focused on antitrust have become more sophisticated in their own area but more disconnected from other work in economics that has considerable potential to improve merger analysis.

Political and practical constraints inevitably limit how much economics can contribute to the social good. But economists interested in antitrust can boost the quantity and quality demanded by shifting out the supply curve ever more. This can be done through continued theoretical and empirical work in industrial organization and increased collaboration with economists in related fields. Some of this research agenda is elaborated in Kaplow (2024b).

■ *I thank Daniel Francis, Nina Pavcnik, Carl Shapiro, Timothy Taylor, Heidi Williams, and the referees for helpful comments and Parth Kotak for research assistance. Disclaimer: I consult on antitrust matters for both government agencies and private parties. In addition, my wife is a lawyer who has mostly represented the financial services industry.*

References

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62 (5): 777–95.

Atkinson, Anthony B., and Joseph E. Stiglitz. 1976. "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics* 6 (1–2): 55–75.

Baker, Jonathan B., and Timothy F. Bresnahan. 1985. "The Gains from Merger or Collusion in Product-Differentiated Industries." *Journal of Industrial Economics* 33 (4): 427–44.

Basu, Susanto. 2019. "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence." *Journal of Economic Perspectives* 33 (3): 3–22.

Bernstein, Lisa. 2015. "Beyond Relational Contracts: Social Capital and Network Governance in Procurement Contracts." *Journal of Legal Analysis* 7 (2): 561–621.

Berry, Steven. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25 (2): 242–62.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.

Berry, Steven, and Ariel Pakes. 1993. "Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis." *American Economic Review* 83 (2): 247–52.

Caradonna, Peter, Nathan Miller, and Gloria Sheu. Forthcoming. "Mergers, Entry, and Consumer Welfare." American Economic Journal: Microeconomics.

Chandler, Alfred D., Jr. 1990. *Scale and Scope: The Dynamics of Industrial Capitalism*. Harvard University Press.

Coase, R. H. 1937. "The Nature of the Firm." *Economica* 4 (16): 386–405.

Cunningham, Colleen, Florian Ederer, and Song Ma. 2021. "Killer Acquisitions." *Journal of Political Economy* 129 (3): 649–702.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics* 135 (2): 561–644.

De Loecker, Jan, and Chad Syverson. 2021. "An Industrial Organization Perspective on Productivity." In *Handbook of Industrial Organization*, Vol. 4, edited by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, 141–223. North-Holland.

Demsetz, Harold. 1973. "Industry Structure, Market Rivalry, and Public Policy." *Journal of Law and Economics* 16 (1): 1–9.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67 (3): 297–308.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62 (1): 53–82.

Farrell, Joseph, and Carl Shapiro. 1990. "Horizontal Mergers: An Equilibrium Analysis." *American Economic Review* 80 (1): 107–26.

Farrell, Joseph, and Carl Shapiro. 2010. "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition." *B. E. Journal of Theoretical Economics* 10 (1): 9.

Fisher, Franklin M. 1987. "Horizontal Mergers: Triage and Treatment." *Journal of Economic Perspectives* 1 (2): 23–40.

Friedman, James W. 1971. "A Non-cooperative Equilibrium for Supergames." *Review of Economic Studies* 38 (1): 1–12.

Gibbons, Robert, and John Roberts. 2013. *The Handbook of Organizational Economics*. Princeton University Press.

Gilbert, Richard J., and Michael L. Katz. 2022. "Dynamic Merger Policy and Pre-Merger Product Choice by an Entrant." *International Journal of Industrial Organization* 81: 102812.

Gowrisankaran, Gautam. 1999. "A Dynamic Model of Endogenous Horizontal Mergers." *RAND Journal of Economics* 30 (1): 56–83.

Green, Edward J., and Robert H. Porter. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica* 52 (1): 87–100.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (4): 691–719.

Hall, Robert E. 2018. "New Evidence on the Markup of Prices over Marginal Costs and the Role of Mega-Firms in the US Economy." NBER Working Paper 24574.

Harrington, Joseph E., Jr. 2017. *The Theory of Collusion and Competition Policy*. MIT Press.

Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98 (6): 1119–58.

Holmstrom, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." *American Economic Review* 84 (4): 972–91.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127–50.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3 (4): 305–60.

Kaplow, Louis. 2011. "On the Optimal Burden of Proof." *Journal of Political Economy* 119 (6): 1104–40.

Kaplow, Louis. 2013. *Competition Policy and Price Fixing*. Princeton University Press.

Kaplow, Louis. 2021. "Market Power and Income Taxation." *American Economic Journal: Economic Policy* 13 (4): 329–54.

Kaplow, Louis. 2023. "Competition Policy in a Simple General Equilibrium Model." *Journal of Political Economy Microeconomics* 1 (1): 80–114.

Kaplow, Louis. 2024a. "Optimal Income Taxation." *Journal of Economic Literature* 62 (2): 637–738.

Kaplow, Louis. 2024b. *Rethinking Merger Analysis*. MIT Press.

Kaplow, Louis, and Carl Shapiro. 2007. "Antitrust." In *Handbook of Law and Economics*, Vol. 2, edited by A. Mitchell Polinsky and Steven Shavell, 1073–1225. North-Holland.

Lancaster, Kelvin. 1980. "Competition and Product Variety." *Journal of Business* 53 (3): S79–103.

Lerner, A. P. 1934. "The Concept of Monopoly and the Measurement of Monopoly Power." *Review of Economic Studies* 1 (3): 157–75.

Lipsey, R. G., and Kelvin Lancaster. 1956. "The General Theory of Second Best." *Review of Economic Studies* 24 (1): 11–32.

Mankiw, N. Gregory, and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *RAND Journal of Economics* 17 (1): 48–58.

March, James G., and Herbert A. Simon. 1958. *Organizations*. Wiley.

Marshall, Robert C., and Leslie M. Marx. 2012. *The Economics of Collusion: Cartels and Bidding Rings*. MIT Press.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Mermelstein, Ben, Volker Nocke, Mark A. Satterthwaite, and Michael D. Whinston. 2020. "Internal versus External Growth in Industries with Scale Economies: A Computational Model of Optimal Merger Policy." *Journal of Political Economy* 128 (1): 301–41.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69 (2): 307–42.

Nocke, Volker, and Michael D. Whinston. 2022. "Concentration Thresholds for Horizontal Mergers." *American Economic Review* 112 (6): 1915–48.

Rasmusen, Eric. 1988. "Entry for Buyout." *Journal of Industrial Economics* 36 (3): 281–99.

Shapiro, Carl, and Ali Yurukoglu. 2024. "Trends in Competition in the United States: What Does the Evidence Show?" NBER Working Paper 32762.

Simon, Herbert A. 1991. "Organizations and Markets." *Journal of Economic Perspectives* 5 (2): 25–44.

Spence, Michael. 1976. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies* 43 (2): 217–35.

Stigler, George J. 1950. "Monopoly and Oligopoly by Merger." *American Economic Review* 40 (2): 23–34.

Stigler, George J. 1964. "A Theory of Oligopoly." *Journal of Political Economy* 72 (1): 44–61.

Stinchcombe, Arthur L. 1990. *Information and Organizations*. University of California Press.

Syverson, Chad. 2019. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives* 33 (3): 23–43.

US Department of Justice and Federal Trade Commission. 2010. *Horizontal Merger Guidelines*. https://www.justice.gov/atr/file/810276/dl?inline.

US Department of Justice and Federal Trade Commission. 2023. *Merger Guidelines*. https://www.ftc.gov/system/files/ftc_gov/pdf/2023_merger_guidelines_final_12.18.2023.pdf.

Werden, Gregory J., and Luke M. Froeb. 1998. "The Entry-Inducing Effects of Horizontal Mergers: An Exploratory Analysis." *Journal of Industrial Economics* 46 (4): 525–43.

Whinston, Michael D. 2007. "Antitrust Policy toward Horizontal Mergers." In *Handbook of Industrial Organization*, Vol. 3, edited by Mark Armstrong and Robert Porter, 2369–2440. North-Holland.
Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press.
Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. Free Press.
Willig, Robert D. 1991. "Merger Analysis, Industrial Organization Theory, and Merger Guidelines." *Brookings Papers on Economic Activity, Microeconomics* 22: 281–312.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 53–76

Acquisitions to Enter New Markets

Carl Shapiro

The Merger Guidelines issued in December 2023 break new ground by announcing that the US Department of Justice and the Federal Trade Commission (2023) will challenge mergers and acquisitions that "could enable the merged firm to extend a dominant position from one market into a related market." In this article, I refer to these as "cross-market acquisitions" or in some cases as "conglomerate mergers." Specifically, the antitrust agencies seek to prevent acquisitions that could "extend a dominant position through exclusionary conduct, weakening competitive constraints, or otherwise harming the competitive process."

This expansion opens a vast new front, because almost all of the mergers reported to the antitrust agencies involve firms that do not compete directly against each other: only 4.6 percent of the deals reported during Fiscal Year 2023 (79 out of 1,735) involved firms that operated even in the same broad three-digit NAICS sector of the economy.[1] Many deals not previously covered by the agencies' merger guidelines now face the risk of an antitrust challenge.

With this major change in merger enforcement, the US antitrust agencies are in danger of putting forward an approach to cross-market acquisitions that echoes in many respects the approach taken 50–60 years ago, which was widely

■ *Carl Shapiro is Professor of the Graduate School, Haas School of Business, University of California at Berkeley, Berkeley, California. His email is cshapiro@berkeley.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20241414.

[1] See Table XI from the Hart-Scott-Rodino Annual Report, Fiscal Year 2023. These three-digit NAICS sectors encompass broad categories such as "Mining (except Oil and Gas)" (212), "Food Manufacturing" (311), and "Telecommunications" (517). These categories include many products that are not substitutes, so the number of reported horizontal mergers must be less than 79, the number of deals involving firms in the same three-digit code.

rejected by scholars, practitioners, and the courts; differs sharply from how the US antitrust agencies have evaluated these acquisitions for at least 40 years; and is inconsistent with how the European Union evaluates them. That approach was abandoned for good reason: acquisitions to enter new markets have long been a central force of economic dynamism, not a category of deals that should be broadly suppressed.

To see what is at stake, consider two examples of cross-market acquisitions, loosely based on real-world examples. First, suppose that a major pharmaceutical company with blockbuster Drug A seeks to acquire Drug B, the leading drug in a different therapeutic category. The stated goal of the acquisition is to deploy the acquiring firm's extensive sales and distribution capabilities to more efficiently market and distribute the acquired Drug B. This acquisition may now be challenged based on the theory that the merged firm might offer bundled discounts to customers who purchase both Drug A and Drug B, thereby placing rivals that only offer Drug B at a competitive disadvantage that will discourage them from investing.

As a second example, suppose that the market leader in operating systems for desktop computers seeks to acquire one of several firms that are developing operating systems for mobile devices. The stated goal of the acquisition is to combine the expertise and intellectual property of the two companies to speed up the development of an innovative new operating system for mobile devices that will disrupt the market for mobile devices. This acquisition may now be challenged based on the theory that it will allow the merged firm to obtain a leading position in the provision of operating systems for mobile devices, causing rival developers of those operating systems to cut back their investments.

In this essay, I begin with brief discussion of the empirical evidence regarding how acquisitions are used to enter new markets. Throughout American economic history, successful firms have diversified and grown their operations by expanding into new markets, including adjacent markets and markets they created with their own innovative new products and services. These expansions have played an important role in fueling economic growth. The challenge for antitrust is to avoid discouraging those deals, while preventing others that harm competition.

I then explain how such acquisitions are treated under the 2023 Merger Guidelines, contrasting that with their prior treatment in the United States and with their treatment in Europe. Ultimately, the 2023 Merger Guidelines are not clear or specific about how the antitrust agencies will analyze cross-market acquisitions that enable firms dominant in one market to enter a new market. This is just one example of their general ambiguity and lack of clarity (as discussed by Francis in this symposium). Thus, I see this portion of the 2023 Merger Guidelines not as offering a finished analysis, but rather as having planted a seed. Prior guidelines have planted some seeds that have grown into mighty trees; for example, the 1992 and 2010 versions of the merger guidelines described what are now widely used methods for assessing whether a horizontal merger is likely to lead to higher prices due to the elimination of price competition between the merging firms.

Will the skepticism about cross-market acquisitions expressed in the 2023 Merger Guidelines primarily promote competition by preventing powerful firms from expanding their empires into new markets? Or will it primarily hinder economic growth by deterring acquisitions that allow successful firms to exploit economies of scale and scope to inject greater competition into new markets—not only those just emerging but also those with entrenched incumbents? I work through how economic evidence and theory can help to distinguish "bad" cross-market acquisitions (those likely to enable the acquiring firm to weaken competition and harm customers in the market it is entering) from "good" ones (those likely to promote competition and benefit customers by enabling a firm that has been successful in one market to more rapidly or more effectively enter a new market). My hope is to offer an instruction manual, complete with health and safety warnings, for nurturing this latest seed so it can grow into a beneficial plant rather than a noxious weed.

Acquisitions to Enter New Markets and Economic Growth

Throughout American economic history, successful firms have diversified and grown their operations by expanding into new markets. These expansions have played a vital role in accelerating the adoption and diffusion of new technologies, in promoting competition, in enhancing efficiency by exploiting economies of scope, and in creating new products and industries. Chandler (1992, p. 96) put it this way: "The ability of large established firms to use learned routines and integrated capabilities to enter related product markets helps to explain a significant change in the ways in which major new industries are coming to be created."

Many acquisitions to enter new markets are neither horizontal nor vertical. They are often referred to as "conglomerate mergers" (for an outstanding, detailed discussion of many aspects of conglomerate mergers, see Church 2008). That genus includes three species, which differ based on whether the products sold by the merging firms are complements, independent on the demand side but purchased by the same group of customers, or unrelated on the demand side.

Cross-market mergers that combine complements create an inherent and beneficial incentive for the merged firm to charge lower prices and provide better products to customers because the merged firm internalizes the profits from all of the complements it owns. Google offers Android free of charge to cell phone manufacturers because Google earns substantial search advertising revenue from those devices. The internalization of profits among complements is typically referred to as "Cournot complements" in honor of Antoine Cournot who identified this effect back in 1838 (for a modern treatment, see Shapiro 2001, pp. 149–50). Cournot complements are common. Moreover, the same economic logic applies to *vertical* mergers, which internalize the profits from the upstream input and the downstream finished good. In that context, internalization is referred to as the "elimination of double marginalization." Mergers involving complements and vertical mergers both tend to benefit customers because they give the merged firm a built-in incentive to lower prices and improve product quality.[2]

Moreover, whether or not the products sold by the merging firms are related on the demand side, firms making acquisitions to enter new markets typically hope and expect to achieve some type of economy of scope on the supply side. In his magisterial studies of the rise of the modern industrial enterprise throughout the twentieth century, Chandler's (1977, 1990) central conclusion is that those enterprises played a fundamental role in bringing about "the most rapid economic growth in the history of mankind" (1990, p. 3). The ability of these firms to diversify their operations and exploit economies of scope by expanding into new areas was critical to their success. Chandler (1990, p. 169) concludes, "Such diversification certainly increased competition within American industry."

The opportunities to exploit economies of scope shifted as the twentieth century progressed. Chandler (1990, p. 188) distinguishes economies of scope within functional operating units from those within the enterprise as a whole. As technology evolved, "economies of scope within the enterprise as a whole provided an even stronger dynamic for growth. One division was able to use intermediate products produced or developed in others, to exploit research and development information and techniques perfected in other divisions, or to apply knowledge acquired in other divisions that used comparable production technologies or served similar markets."

Similar patterns have continued into the twenty-first century. For example, today's leaders in cloud computing—Amazon, Google, and Microsoft—launched this rapidly growing market by exploiting economies of scope with their other lines of business. As another example, Amazon and Apple exploited economies of scope to enter the market for video streaming services, which had been inhabited by powerful incumbents. Lee and Lieberman (2024) describe the diverse ways in which Amazon and Alphabet have used internal growth versus acquisition to enter new lines of business within and beyond their primary domains.

Firms seeking to expand into new areas often compare internal growth with acquisition. Both methods are common—and they are not as distinct as one might think. For example, when a large pharmaceutical company purchases a biotech startup with a promising drug in the pipeline, it may also consider the alternative choice of licensing the intellectual property rights to that drug for its own internal development. More generally, expansion by "internal growth" typically involves the purchase of various assets, and the US merger statutes have covered asset purchases as well as stock purchases since 1950. Even if one restricts attention to stock purchases, so-called "foothold acquisitions" are a common intermediate case.

[2] To gain some intuition on the isomorphism between vertical mergers and mergers involving complements, consider a vertical merger between Firm A, which sells Input A, and Firm B, which sells finished Good B to Customer C. If Firm B has been buying Input A from Firm A, this is a vertical merger. However, if Customer C instead has been buying Input A itself directly from Firm A and combining it with Good B to produce the finished good, this same merger combines complements. Both patterns arose in the GE/Honeywell case discussed below.

These transactions arise when a large firm acquires a relatively small firm, perhaps a startup, in another market and then makes substantial investments to grow its business in that market. Foothold acquisitions can disrupt the target market and threaten incumbents in that market. Google's acquisitions of YouTube and Android are good examples of foothold acquisitions.

The choice between expanding via internal growth versus acquisition has much in common with firms' traditional "make versus buy" decisions, which have been extensively analyzed, most famously by Williamson (1975). Expanding into a new market via an acquisition is typically more expensive than doing so via internal growth, but also faster and less risky. If a firm prioritizes speed and lower risk, the most effective way to enter a new market will often involve acquiring an established firm, either one already operating in the new market or one with the components or capabilities required to compete in that market.

Chandler (1990) provides numerous historical examples of what he calls "diversification through merger," by which he means modern industrial enterprises that achieved economies of scope by acquiring existing firms in new markets. In the food and chemicals sector, most of the large enterprises producing branded packaged products "combined large integrated firms whose complementary product lines permitted a more intensive use of their joint facilities and skills" (p. 165). For example, several firms acquired new products so they could better utilize their own distribution networks that delivered perishable goods on a daily basis. In other cases, acquisitions enabled firms to leverage their organizational capabilities to enter new markets.

A prominent early study argued that firms tend to enter new markets via internal growth if synergies are likely based on proximity, whereas they tend to enter new markets via acquisition if entry barriers are high (Yip 1982). The subsequent literature is more nuanced. For example, in the telecommunications industry between 1989 and 2003, Lee and Lieberman (2010) find that "inside a firm's primary business domain, acquisitions are used to fill persistent gaps near the firm's existing products, whereas outside that domain, acquisitions are used to extend the enterprise in new directions."

Empirical studies have found a variety of effects associated with cross-market acquisitions in the manufacturing sector made by publicly-traded companies during the late twentieth century. Such acquisitions can promote the efficient intra-firm transfer of intangible inputs (Atalay, Hortaçsu, and Syverson 2014; Bet 2024) and improve the productivity of the acquired facility (Schoar 2002; Bet 2024). However, they also may reduce the productivity of previously owned facilities by diverting management attention away from them, in what Schoar (2002) calls a "new toy" effect. One should not be surprised that some of these acquisitions enhance efficiency and are successful while others are failures, some of which are later unwound. Variety is the norm in business, and experimentation often promotes competition and economic growth.

In the twenty-first century, acquisitions remain an important way for firms to enter new markets and exploit economies of scope, as acquiring firms seek synergies by combining their own capabilities with the assets they are acquiring. This dynamic is of great practical importance because firms differ very widely and persistently in their efficiency, as explained and documented in convincing detail by Bloom and Van Reenen (2007, 2010), Syverson (2011), and De Loecker and Syverson (2021). Such acquisitions can enable the assets of the acquired firm to be utilized more efficiently, and they can allow successful and innovative firms to challenge powerful incumbents by entering new markets. There are good reasons, both theoretically and empirically, to expect that internal growth often will be inferior to acquisition in serving both of these functions. In the tech sector, the majority of acquisitions during 2010–2020 fell outside the acquiring company's core area of business (Jin, Leccese, and Wagman 2023).

In short, many acquisitions to enter new markets over the past century have been beneficial, and many future acquisitions of this type are likely to promote competition and economic growth. But what about the exceptions: acquisitions that harm customers by leading to a monopoly in the newly entered market? Chandler (1990, p. 189) did find that "occasionally" the firm diversifying into a new market would obtain a near monopoly, "such as Du Pont acquired in moisture-proof cellophane, nylon, and tetraethyllead." In some of those cases, such as cellophane, the firm *created* the new market by innovating. In any event, Chandler found that more often multiple firms "responded to the same market opportunity or quickly followed the pioneer into the new product market." Chandler thus concluded that these "new strategies of diversification, then, not only intensified competition in existing industries but helped ensure competition in new ones."

So far as I am aware, specific examples of proven deleterious cross-market mergers are few and far between. Identifying such cases and their characteristics is difficult but would be very valuable. As one prominent category of candidates, several recent studies have shown that cross-market mergers involving hospitals in different geographic areas have led to higher prices.[3] In these cases, the merging hospitals were too far apart to be substitutes for patients. However, they may have been substitutes from the perspective of the health insurers that negotiate with hospitals for inclusion in their covered networks. Some of these deals may therefore actually have been horizontal mergers.

This evidence highlights an important lesson for antitrust rules about acquisitions to enter new markets: as the antitrust agencies raise the costs and risks of expanding into new markets via acquisition, would-be acquiring firms will either

[3] Lewis and Pflum (2017) show that the prices at hospitals acquired by out-of-market hospital systems during 2000–2010 rose by about 17 percent more than unacquired, stand-alone hospitals. They attribute these higher prices to increased bargaining power for the acquired hospitals in their negotiations with managed care organizations. Similarly, Dafny, Ho, and Lee (2019) find that cross-market, within-state hospital mergers during the 1996–2012 period raised prices by 7 percent to 9 percent, while similar out-of-state acquisitions did not lead to significant price increases. Arnold et al. (2024) obtain similar results for the 2009–2017 time period. Peters (2014) shows in a bargaining model how cross-market mergers can raise prices. Vistnes (2024) provides a thoughtful, recent antitrust analysis of cross-market hospital mergers.

scale back their expansion plans or resort instead to internal growth, even in cases where that route is slower or less likely to succeed. In many cases, that would put less pressure on incumbents in the target market, thereby harming customers and weakening competition.

The challenge is to identify acquisitions to enter new markets that do *not* promote competition so interventions can be targeted at them. Do the 2023 Merger Guidelines explain how the antitrust agencies will do that?

Acquisitions to Enter New Markets under the 2023 Merger Guidelines

Section 2.6.B of the 2023 Merger Guidelines is entitled "Extending a Dominant Position into Another Market." That section begins:

> The Agencies also examine the risk that a merger could enable the merged firm to extend a dominant position from one market into a related market, thereby substantially lessening competition or tending to create a monopoly in the related market. For example, the merger might lead the merged firm to leverage its position by tying, bundling, conditioning, or otherwise linking sales of two products.

I refer to the market in which the acquiring firm has a dominant position as the "core" market, and the market in which the acquired firm operates as the "new" market.

As authority for its policy toward acquisitions to enter new markets, the 2023 Merger Guidelines cite a 1972 Supreme Court case involving Ford Motor Company's 1961 acquisition of Autolite, a manufacturer of spark plugs (*Ford Motor Co. v. United States*, 405 US 562 [1972]). However, the Ford/Autolite case is uninformative regarding the evaluation of mergers that are neither horizontal nor vertical, because (as the Supreme Court recognized) it was *both*. Ford's acquisition of Autolite was *horizontal* because Ford was a potential entrant into the market for spark plugs. Ford's threat of entry into spark plugs was found to have disciplined the existing suppliers of spark plugs. The loss of Ford as a potential entrant into spark plugs provided one reason why the Supreme Court found Ford's acquisition of Autolite to be illegal. This is a traditional horizontal (potential competition) theory of harm. Ford's acquisition of Autolite also was *vertical* because Ford was the largest purchaser of spark plugs from independent suppliers. The loss of Ford as a potential customer for independent suppliers of spark plugs provided the second reason why the Supreme Court found the acquisition to be illegal. This is a traditional vertical (customer foreclosure) theory of harm.

Moreover, the single nonhorizontal, nonvertical acquisition cited in the 2023 Merger Guidelines is worrisome. That case involved Procter & Gamble's 1957 acquisition of Clorox (*Federal Trade Commission v. Procter & Gamble Co.,* 386

US 568 [1967]). Clorox was the leading manufacturer of household liquid bleach, with a market share of 49 percent, and Procter & Gamble was a large, diversified manufacturer of other household products. The Supreme Court ruled in 1967 that this acquisition had harmed competition in the market for household liquid bleach because "the substitution of the powerful acquiring firm [Procter & Gamble] for the smaller, but already dominant, firm [Clorox] may substantially reduce the competitive structure of the industry by raising entry barriers and by dissuading the smaller firms from aggressively competing." The Court was specifically concerned that "Procter would be able to use its volume discounts to advantage in advertising Clorox. Thus, a new entrant would be much more reluctant to face the giant Procter than it would have been to face the smaller Clorox." In other words, the merger was found to be illegal because it lowered the costs of the market leader, Clorox, making it an even more formidable competitor. This acquisition may well have harmed Clorox's rivals, but it does not appear to have harmed consumers who purchased household liquid bleach.

The *Procter & Gamble* decision was representative of the jurisprudence of the 1960s, but that jurisprudence shifted sharply during the 1970s. Bauer (1978) reports that were eleven successful challenges to conglomerate mergers between 1964 and 1974. He also reports that none of the twelve challenges to conglomerate mergers between 1974 and 1977 were successful. Bauer (1983) updates this analysis, reporting (p. 350) that "the continued failure of court challenges against conglomerate mergers [after 1977] has been dramatic."

As we seek to understand the new policy toward cross-market acquisitions in the 2023 Merger Guidelines, three key questions arise. First, which acquiring firms will be seen by the antitrust agencies as possessing a "dominant position" in their core market, so their acquisitions might run afoul of the new policy? Second, what potential post-merger conduct will be regarded as "exclusionary"? Third, which cross-market acquisitions will be challenged based on the *horizontal* theory that the acquiring firm is a potential entrant into the new market?

What Constitutes a "Dominant Position"?

The 2023 Merger Guidelines state that the antitrust agencies will assess whether a firm has a dominant position "based on direct evidence or market shares showing durable market power." They do not provide concrete examples or market share thresholds. However, the Guidelines do cite the *Ford/Autolite* case, in which Ford was the second largest manufacturer of automobiles (General Motors was the leader) with a share of about 30 percent. That strongly suggests that the enforcement agencies will view a share of 30 percent in the core market as sufficient to establish dominance, even if another firm has a larger share.[4]

[4] The early draft of the Merger Guidelines released for public comment in June 2023 (available at https://www.ftc.gov/system/files/ftc_gov/pdf/p859910draftmergerguidelines2023.pdf) explicitly stated that a firm with a market share of at least 30 percent had a dominant position, but that language was not retained in the final version. Ultimately, the question is whether the *courts* will accept a market share as low as

The 2023 Merger Guidelines apply the same approach to dominance in the new market: an acquisition may be seen as extending a dominant position into the new market even if there is no prospect that the merged firm will obtain a majority share of that market. They explicitly refer to deals that allow the acquiring firm "to obtain a foothold in another market." Notably, in the *Ford/Autolite* case, Autolite's market share of spark plugs when it was acquired by Ford was only 15 percent, the leading suppliers of spark plugs were Champion, with 50 percent of the market, followed by General Motors (with its AC brand) with 30 percent, and there was no real prospect that Ford would capture more than about 30 percent of the market for spark plugs.[5]

The 2023 Guidelines do not differentiate between the merged firm gaining a dominant position in an existing market versus a newly created one. Thus, the text suggests that an acquisition may be seen extending a dominant position into a new market even if the merged firm *creates* the new market by offering a novel product or service. Likewise, the 2023 Guidelines can be read to suggest that if disruption of the new market will be so substantial as to make the merged firm a major competitive force, that success itself might make the merger illegal.

What Post-Acquisition Conduct Will be Regarded as Exclusionary?

The 2023 Merger Guidelines say very little about the types of evidence that will trigger a challenge based on the fear that the merged firm will engage in exclusionary conduct. "Exclusionary conduct" is not defined, and the Guidelines do not say if or how the antitrust agencies will distinguish between acquisitions that harm *rivals* in the new market and those that harm *competition* in that market. That is unfortunate, because any acquisition that injects competition into the new market can be expected to harm rivals, possibly even causing them to exit. The reader cannot tell if or how the 2023 Merger Guidelines adopt and apply the core principle that antitrust law protects competition, not competitors (as discussed by Francis in this symposium).

Here is the entirety of the guidance found in Section 2.6.B of the 2023 Merger Guidelines:

> For example, the merger might lead the merged firm to leverage its position by tying, bundling, conditioning, or otherwise linking sales of two products. A merger may also raise barriers to entry or competition in the related market, or eliminate a nascent competitive threat as describe above. For example, prior to a merger, a related market may be characterized by scale economies but still experience moderate levels of competition. If the merged firm takes

30 percent as evidence of a dominant position. In the past, significantly higher shares have been needed to establish monopoly power.

[5] The Supreme Court was concerned that Ford would shift its purchases of spark plugs from Champion and other independents to Autolite—even though this would *reduce* concentration in the market for spark plugs. Indeed, by 1966, five years after the merger, Champion's share had fallen from 50 percent to 33 percent.

> actions to induce customers of the dominant firm's product to also buy the related product from the merged firm, the merged firm may be able to gain dominance in the related market, which may be supported by increased barriers to entry or competition that result from the merger.

The Guidelines do not indicate whether "tying, bundling, conditioning, or otherwise linking sales of two products" will always be seen as exclusionary. Nor do they identify circumstances under which such conduct will *not* be seen as exclusionary. In any event, a wide variety of conduct is covered by the guidance given above: what merged firm would *not* try to induce its established customers to also buy the related product from it?

The general idea seems to be that the merged firm might adopt tactics to induce customers who purchase its dominant product to also purchase the product it has acquired in the new market. The central concern is that by using such tactics the merged firm will achieve at least a 30 percent share in the new market and thereby weaken its rivals in that market by depriving them of scale economies.

The 2023 Merger Guidelines warn: "A merger can result in durable market power and long-term harm to competition even when it initially provides short-term benefits to some market participants." This language suggests that the antitrust agencies may challenge a cross-market acquisition that provides near-term benefits to customers in the new market based on the fear that the acquisition might weaken rivals in that market over time by taking business away from them, thus reducing their incentive to make investments. The Guidelines implicitly discount, without mentioning, the possibility that rivals will instead redouble their efforts in response to increased competitive pressure by offering better deals to customers and investing more. Moreover, they do not address a central reason why cross-market acquisitions take business from rivals in the new market—because the acquired firm offers improved products or lower prices, just the sort of competitive pressure the antitrust laws seek to encourage.

How will how the antitrust agencies determine what conduct (exclusionary or not) is likely to occur after the merger? The traditional approach taken by the antitrust agencies when assessing post-merger conduct for horizontal and vertical mergers is to focus on conduct that the merged firm will have both the ability and the incentive to pursue. In some cases, this sensible and workable "ability and incentive" approach will rely on specific evidence about the merged firm's post-merger plans. In others, it will be based on economic expert analysis informed by the acquiring firm's pre-merger conduct and market realities.

Will the Acquisition Be Seen as Eliminating a Potential Entrant?

Blocking mergers that eliminate future competition is very important. In Rose and Shapiro (2022), we suggested updating the 2010 Horizontal Merger Guidelines to say more about dynamic competition, including acquisitions of potential competitors. However, the 2023 Guidelines express an unqualified preference for entry via internal growth rather than via acquisition. Guideline 4, which involves acquisitions of potential competitors, states: "In general, expansion into a concentrated market via internal growth rather than via acquisition benefits competition. Merging a current and a potential market participant eliminates the possibility that the potential entrant would have entered on its own—entry that, had it occurred, would have provided a new source of competition in a concentrated market." That view is unsupported by economic evidence or theory, as elaborated above.

The antitrust agencies will challenge a cross-market acquisition in a concentrated market based on the loss of potential competition if two prongs are satisfied: (1) the acquiring firm "had a reasonable probability of entering" the target firm's market without the acquisition; and (2) entry in that manner "offered a substantial likelihood of ultimately producing deconcentration."[6]

To see some of the pitfalls here, consider the common situation in which the acquiring firm is seeking enter a new market to take advantage of economies of scope with its core operations. Suppose the new market is "concentrated" as defined by the 2023 Merger Guidelines, meaning that the Herfindahl-Hirschman Index is greater than 1000. For the first prong, the antitrust agencies look at "objective evidence regarding the firm's available feasible means of entry, including its capabilities and incentives." They also state that "[s]ubjective evidence that the company considered organic entry as an alternative to merging generally suggests that, absent the merger, entry would be reasonably probable." If the firm studied organic entry but rejected it as infeasible, or simply too slow and risky to justify the required investment, the first prong will evidently be satisfied. That is baffling. The second prong will be satisfied if the acquiring firm would "ultimately" have gained any meaningful market share in the new market.

The bottom line is that many cross-market mergers now face the risk of being challenged by the antitrust agencies, either because the acquiring firm will be seen as a potential entrant into the new market or because they will be seen as extending a dominant position from the firm's core market into a new market (or both). The latter risk is accentuated because the 2023 Merger Guidelines define a "dominant position" broadly and because they provide so little guidance about what post-merger conduct will be regarded as exclusionary.

Case Study: Amgen's Acquisition of Horizon Therapeutics

For an informative case study of how the antitrust agencies are currently approaching cross-market acquisitions, consider the 2023 challenge by the Federal Trade Commission to Amgen's $27.8 billion acquisition of Horizon Therapeutics. (The FTC Complaint is here: https://www.ftc.gov/system/files/ftc_gov/pdf/2310037amgenhorizoncomplainttropi.pdf.)

[6] The discussion here follows Section 2.4.A of the 2023 Merger Guidelines, which explains how the antitrust agencies will determine whether the acquiring firm is an "actual potential entrant" into the target market. Section 2.4.B states that they also may challenge a transaction if they conclude that the acquiring firm is a "perceived potential entrant." They may reach that conclusion if another market participant could reasonably *consider* the firm to be a potential entrant, even if the firm itself decided against entry via internal growth.

The Federal Trade Commission asserted that Amgen had a practice of providing greater rebates to its customers (pharmaceutical benefit managers, health plans, and plan sponsors) on its blockbuster drugs in exchange for their giving favorable formulary placement to other Amgen medications in unrelated product markets. The FTC alleged that Amgen would offer rebates to its customers on its blockbuster drugs in exchange for their providing favorable treatment to two specific Horizon drugs, which would suppress emerging competition for those drugs. The FTC alleged: "Cross-market rebating and bundling can also block smaller rivals from being able to compete on the merits. . . . If permitted to acquire Horizon, Amgen would have the ability and incentive to sustain and entrench the monopolies of Horizon's drugs using similar multi-product contracting strategies." The FTC Complaint did not point to any specific evidence that Amgen actually planned to offer such cross-market rebates.

The Federal Trade Commission focused on how Amgen's possible cross-market rebates would reduce the incentives of actual and potential rivals to develop drugs that compete against Horizon's drugs, not on how these rebates would affect *patients*. The FTC thus appeared to view any near-term benefits that customers would get from such rebates as outweighed by the harms associated with the possible perpetuation of Horizon's monopoly over those two drugs. To add a final twist, Amgen (2023) promised not to offer bundles involving the Horizon drugs. The FTC initially declined that offer and sued Amgen, but ultimately accepted it as a remedy to settle its case against Amgen (Federal Trade Commission 2023).

Earlier US and EU Policies toward Cross-Market Acquisitions

For decades, the US antitrust agencies have not challenged any nonhorizontal, nonvertical mergers based on the theory that the merged firm will extend a dominant position from its core market into a new market by linking the sales of the acquired product with those of its core product. However, they did challenge some conglomerate mergers back in the 1960s and 1970s. Now that those theories have been revived in the 2023 Merger Guidelines, it is important to understand why they were abandoned decades ago.

The Evolution of US Merger Guidelines on Conglomerate Mergers

Between 1965 and 1975, the United States experienced a wave of conglomerate mergers. On the whole, those mergers were unsuccessful strictly from a *business* perspective (Ravenscraft and Scherer 1987, 1989). But our interest here is to understand how the antitrust agencies and the courts responded to that passing fad of conglomerate mergers.

The first Merger Guidelines, issued in 1968, provide a good starting point. They state that the US Department of Justice may challenge "a merger of firms producing related products which may induce purchasers, concerned about the merged firm's possible use of leverage, to buy products of the merged firm rather than those of competitors." The term "leverage" was never defined in the 1968 Guidelines, but it likely referred to what the 2023 Merger Guidelines call "linking sales of two products." The 1968 Guidelines explain why they are not more specific: "Generally speaking, the conglomerate merger area involves novel problems that have not yet been subjected to as extensive or sustained analysis as those presented by horizontal and vertical mergers. It is for this reason that the Department's enforcement policy regarding the foregoing category of conglomerate mergers cannot be set forth with greater specificity."

The late 1960s was a time of considerable hostility toward all acquisitions by large firms. Coincident with the 1968 Merger Guidelines, a Task Force on Antitrust Policy appointed by President Lyndon Johnson recommended a new merger act "that would prohibit any 'large firm' from acquiring or merging with any 'leading' firm." The Task Force gave this rationale: "If large firms are prevented from acquiring leading firms in concentrated industries, they will seek other outlets for expansion which may be more likely to increase competition and decrease concentration." The Task Force defined a "large firm" as one with assets of at least $250 million or sales of at least $500 million. A "leading firm" was defined as one with a market share of at least 10 percent (and among the four largest firms) in a market with sales of at least $100 million and with a four-firm concentration ratio of at least 50 percent (Neal et al. 1968, p. 13).[7] This proposal did not move forward.

During the 1960s and 1970s, the antitrust agencies challenged a number of conglomerate mergers, but this line of merger enforcement soon came to be widely seen as poorly conceived (for example, Areeda and Turner 1978). The fundamental problem was that the government was challenging mergers based principally on the harm they might cause to *rivals*, not to customers. The poster child for this approach is the 1967 *Procter & Gamble* case described above. In any event, the antitrust authorities suffered a string of defeats in conglomerate merger cases, especially after 1974 (Bauer 1983).

Regarding the species of conglomerate mergers that combines complements, we can learn a lot from how the agencies and the courts have treated *vertical* mergers. As noted above, the same economic logic by which vertical mergers tend to lead to lower prices and improved products applies to mergers that combine complements. The 2020 Vertical Merger Guidelines emphasize the "elimination of double marginalization" as a pro-competitive benefit commonly associated with vertical mergers (US Department of Justice and Federal Trade Commission 2020). The effect is automatic but may not be merger-specific. Shapiro (2021) provides a real-world application of the elimination of double marginalization and a discussion of how to determine whether it is merger-specific.

Prior to the 2023 Merger Guidelines, the most recent policy statement by the US Department of Justice and the Federal Trade Commission specifically about

[7] Yip (1982, p 344) questioned this rationale based on his empirical findings about entry by internal growth versus acquisition. "Current policy may merely preclude the enhanced market competitiveness sometimes caused by acquisition entry, without gaining new competitors via direct entry."

conglomerate mergers and cross-market acquisitions was United States (2020), a submission by the US government to the OECD. That submission is stunningly different from the 2023 Merger Guidelines, stating: "Conglomerate mergers that raise neither vertical nor horizontal concerns are unlikely to be problematic under U.S. merger law." This submission also emphasizes that conglomerate mergers can enable economies of scope:

> Mergers that combine unrelated products can result in procompetitive benefits if their production or distribution uses the same assets, inputs, or know-how. For instance, when suppliers combine their assets to jointly produce multiple final products for customers, a merger can eliminate contracting frictions and allow for profit maximization over a larger set of products. A single firm able to coordinate how these assets are used may be able to streamline production, inventory management, or distribution . . . A key lesson learned from prior U.S. experience is that, in the absence of evidence of consumer harm in a relevant market, the presence of these sorts of efficiencies benefits competition and consumers, even if the merged firm will become a more effective competitor or gain share.

In 2020, the antitrust agencies noted that they had not challenged any conglomerate mergers in decades. "The Agencies have not brought in modern times any challenges to mergers of unrelated products that rely on 'conglomerate' theories outside the horizontal and vertical frameworks." While the absence of these kinds of enforcement actions could in theory represent a decades-long, bipartisan blind spot, a more plausible hypothesis is that it reflects a well-developed understanding based on economic evidence that cross-market acquisitions often promote competition and only rarely have been shown to harm competition.

Acquisitions to Enter New Markets under the European Merger Guidelines

The European Union has its own guidelines regarding cross-market acquisitions: the Non-Horizontal Merger Guidelines issued by the European Commission in 2008. Like the 2023 US Merger Guidelines, the 2008 EC Guidelines are primarily concerned that the merged firm will harm competition by engaging in tying or bundling to gain business in the new market. Like the 2023 Merger Guidelines, the 2008 EC Guidelines ask whether the merged firm will have the ability and incentive to engage in such tying or bundling.

Despite these similarities, the 2008 EC Guidelines are far less skeptical of cross-market acquisitions than are the 2023 US Merger Guidelines, as illustrated by these passages:

> [I]t is acknowledged that conglomerate mergers in the majority of circumstances will not lead to any competition problems . . . Tying and bundling as such are common practices that often have no anticompetitive consequences. Companies engage in tying and bundling in order to provide their customers

> with better products or offerings in cost-effective ways. . . . Customers may have a strong incentive to buy the range of products concerned from a single source (one-stop-shopping) rather than from many suppliers, e.g. because it saves on transaction costs. The fact that the merged entity will have a broad range or portfolio of products does not, as such, raise competition concerns.

Critically, the 2008 EC Guidelines ask "whether a foreclosure strategy would have a significant detrimental effect on competition, thus causing harm to consumers." The need to show harm to consumers is thus an explicit part of the EC Guidelines, while the 2023 US Merger Guidelines state that the antitrust agencies do *not* require near-term harm to customers. More generally, they leave ambiguous whether harm to customers is necessary at all.

Furthermore, the 2008 EC Guidelines apply well-established economic principles to provide far more specific guidance about how cross-market acquisitions will be analyzed than do the 2023 US Merger Guidelines:

> The effects of bundling or tying can only be expected to be substantial when at least one of the merging parties' products is viewed by many customers as particularly important and there are few relevant alternatives for that product.

> The more customers tend to buy both products (instead of only one of the products), the more demand for the individual products may be affected through bundling and tying.

> [T]he scope for foreclosure tends to be smaller where the merging parties cannot commit to making their tying or bundling strategy a lasting one, for example through technical tying or bundling which is costly to reverse.

> [T]he Commission considers, on the basis of the information available, whether there are effective and timely counter-strategies that the rival firms may deploy. One such example is when a strategy of bundling would be defeated by single-product companies combining their offers so as to make them more attractive to customers.

> In addition, rivals may decide to price more aggressively to maintain market share, mitigating the effect of foreclosure.

The 2008 EC Guidelines also helpfully discuss several real-world examples of cross-market acquisitions that were investigated by the European Commission, including some in which no enforcement action was taken. In a highly visible case, General Electric (GE), the leading supplier of jet engines for large commercial aircraft, sought to acquire Honeywell, a major supplier of avionics equipment for those aircraft. The European Commission (2001) blocked that acquisition based on its concern that GE would extend its dominant position in jet engines into avionics equipment.[8] This caused a rift between the European Commission and the United States, which had not been concerned about GE adding avionics to its line of products and took umbrage at the EC intervention (Kolasky 2002). On appeal, the European Court of First Instance (2005) ruled that the European Commission had not proven that the acquisition would harm competition by extending GE's dominant position from jet engines into avionics.

In recent years, the European Commission has investigated a number of cross-market acquisitions to determine whether they would enable the merged firm to extend a dominant position from one market into another. These cases provide rich detail on how the 2008 EC Guidelines are applied. For example, when Essilor sought to acquire Luxottica, the EC investigated whether that acquisition would allow Essilor to extend its position in the sale of eyeglass lenses into the sale of frames and sunglasses, and whether it would allow Essilor to extend its position in the sale of ophthalmic machines into the sale of eyewear. Ultimately, the European Commission (2018) did not challenge that acquisition.

The 2008 EC Guidelines state that "conglomerate mergers in the majority of circumstances will not lead to any competition problems." They also spell out specific factors that will be considered when evaluating such mergers and offer concrete examples of how these standards will be applied. These features reduce the risk that pro-competitive mergers will be deterred due to antitrust risk, and they allow companies considering such mergers to better assess whether they are likely to be challenged by the antitrust authorities. The 2023 US Merger Guidelines do not share these features.

Economic Principles to Guide the Evaluation of Cross-Market Acquisitions

The 2023 US Merger Guidelines as written are potentially a major step backward regarding the treatment of cross-market acquisitions. Two major problems stand out: ambiguity and lack of balance. Ambiguity arises because they lack specificity and analytical rigor regarding how the antitrust agencies will analyze cross-market acquisitions. Lack of balance arises because the 2023 Merger Guidelines never acknowledge that cross-market acquisitions often benefit customers in the market the acquiring firm is entering by enabling economies of scope. As a result, the US antitrust agencies are now threatening to challenge cross-market acquisitions that would provide near-term benefits to customers in the new market based on a concern that the merged firm will take business from rivals in that market by linking sales of the two products. That stance risks blocking mergers that would inject more competition into the new market.

[8] Patterson and Shapiro (2001) summarize the issues involved and the sources of the trans-Atlantic divergence. I consulted for General Electric in this case.

These problems can be addressed if US antitrust agencies flesh out the relatively brief language found in the 2023 Merger Guidelines. That could be done through speeches, enforcement actions, competitive advocacy, or explicit updates. The key is to ground the treatment of cross-market acquisitions in what we know about their effects, both empirically and theoretically. New modes of analysis announced in prior guidelines have been most influential and durable when they reflect robust advances in economic learning based on real-world cases and research findings.

Evaluation Should Be Based on How Customers Are Affected, Not Rivals

The starting point is to be clear about the goal—to identify and block mergers that harm *customers* as a result of diminished competitive constraints, not those that harm competitors. This distinction is fundamental: after all, if the merged firm offers greater value to customers, rivals will typically be harmed. If protecting rivals is the focus, many acquisitions that would benefit customers by injecting more competition into the new market will be blocked.

To crystallize the issue, consider the acquisition of a perishable branded product by a firm that has innovated by developing a logistical network capable of delivering many perishable products at scale on a daily basis. Suppose that other brands of that product are likely to lose substantial sales after the acquisition because they are currently only capable of making deliveries every few days. Moreover, the acquisition arguably raises entry barriers into the sale of this perishable product, because the bar for success has been raised. Under the 2023 Merger Guidelines, these facts could be seen as sufficient to condemn the merger as illegal.

More generally, consider an acquisition that enables the merged firm to exploit economies of scope, making its costs of providing the product in the new market lower than those of the acquired firm. Suppose the merged firm would pass through some portion of those lower costs to customers in the form of lower prices for the new product by offering multi-product bundles. The direct effect of this acquisition would be to enhance competition and benefit customers in the new market. Rivals in the new market would be harmed by the increased competition (in the form of lower prices) coming from the merged firm. They might well lose sales, making it harder for them to cover their fixed costs, including research and development costs. In this sense, the multi-product bundles may be seen as "exclusionary." Based on this fact pattern alone, blocking this transaction on the ground that it "may substantially lessen competition" in the new market because it harms rivals in that market would be perverse.

While it might seem odd that an acquisition could be illegal because it allows the merged firm to exploit economies of scope to better serve customers, that was precisely the outcome in the 1967 *Procter & Gamble* case, which is discussed above and is cited by the 2023 Merger Guidelines. Unfortunately, the 2023 Guidelines are not clear on this issue. As noted earlier, their focus on rivals, together with the absence of any clear statement that the ultimate concern is about harm to customers, is troubling. The US antitrust agencies could readily issue a statement that a cross-market acquisition will only be challenged if it is expected to harm *customers* as a result of diminished competitive constraints. That would solve this particular problem and focus the analysis. The 2008 EC Guidelines and United States (2020) are both clear on this key point.

Many Acquisitions to Enter New Markets Increase Competition

Economies of scope have been a powerful force in the American economy for more than a century. Any reasonable policy needs to take care not to impede cross-market acquisitions that enable acquiring firms to extend their skills and capabilities into new markets by combining complementary assets and products within a single firm. The prospect that the merged firm will gain considerable share in the new market should in general be seen as a *plus*, not a minus, and especially so if the acquisition is likely to disrupt a market that is initially highly concentrated. Likewise, the prospect that the merged firm will gain a dominant position by *creating* a new market should be seen as a plus, not a minus, even if that dominance might last for some time.

The 2023 Merger Guidelines do not mention that a cross-market merger that combines complements creates an inherent incentive for the merged firm to lower prices and/or increase product quality. That omission is stunning, given the great emphasis that the Guidelines place on how bringing *substitutes* under common ownership creates an inherent and harmful incentive for the merged firm to raise prices and/or reduce product quality. The economic logic of combining complements within the same firm is precisely the same, but with the opposite sign.[9]

Efficiencies in Cross-Market Acquisitions

The antitrust agencies and the courts have long been highly skeptical of efficiency claims put forward by major rivals seeking to merge. Such skepticism makes good sense. In virtually any model of oligopoly, a horizontal merger will increase the joint profits of the merging firms, even without any cost savings or other synergies. Moreover, when the agencies challenge a merger between two major rivals, the inherent anticompetitive effects are likely to be substantial, so it would take significant efficiencies to counter them (Nocke and Whinston 2022). Plus, if economies of scale are sizeable, they often can be achieved as firms compete to expand through internal growth, yielding a better outcome for customers than a merger that short-circuits that process. For all of these reasons, the agencies and the courts have for decades imposed strict requirements on efficiency claims in horizontal mergers: to redeem a merger, they must be merger-specific, carefully verified, and of a character and magnitude sufficient to prevent a lessening of competition.

[9] As a separate point, the case of substitutes diverges from the case of complements when one considers whether the effects are merger-specific. In the case of substitutes, without the merger it would be *per se* illegal under antitrust law for the two merging firms (rivals) to agree to raise their prices. In the case of complements, without the merger there would be no comparable prohibition preventing the two merging firms (not rivals) from agreeing to offer a lower price for a customer purchasing both products. Whether internalization is merger-specific is a key question in vertical and cross-market mergers. Shapiro (2021) discusses this issue in depth for vertical mergers.

The same degree of skepticism is not warranted for efficiency claims associated with cross-market acquisitions. As emphasized above, an acquisition that combines complements gives the merged firm an inherent incentive to provide better value to customers—the opposite of the inherent incentive associated with a horizontal merger. Moreover, the historical record shows that acquisitions to enter new markets often promote competition and spur economic growth. Yet the 2023 Merger Guidelines make no distinction between how efficiency claims associated with horizontal mergers are assessed and how those associated with cross-market acquisitions are assessed. No reason is given for treating these two very different creatures identically, an approach that fails to reflect the historical evidence and is at odds with economic theory. Fixing this error by explaining specifically how efficiencies associated with cross-market acquisitions will be treated would reduce the danger of deterring beneficial transactions of that type.

The Effects of Tying and Bundling Are Complex and Fact-Specific

Under the 2023 Merger Guidelines, the antitrust agencies will examine the risk that a cross-market acquisition "might lead the merged firm to leverage its position by tying, bundling, conditioning, or otherwise linking the sales of the two products." No other specific tactics are identified, and no guidance is given about the market conditions more or less likely to trigger a challenge from the antitrust enforcement agencies. This lack of guidance is unfortunate, especially given the useful guidance provided 15 years earlier in the 2008 EC Guidelines.

Ideally, merger enforcers would identify specific market features that tend to indicate that the tying and bundling made possible by an acquisition are likely to harm customers by weakening competition in the new market rather than benefit them due to cost savings or beneficial incentive effects. That goal is challenging. There have been many antitrust cases involving tying and bundling, but determining how those practices affect competition and customers has proven difficult, even for longstanding practices. Assessing the likely effects on competition and customers of *possible future* tying and bundling is even more difficult.

Economists do have a number of well-crafted theories about the incentive to engage in tying and bundling. These theories offer predictions about the effects of these practices in different settings. We understand that bundling and tying can be profitable based on a number of distinct mechanisms, including offering efficient packages (technical or economic) to customers, saving on transaction costs, assuring product quality, enhancing price discrimination, and excluding rivals. Reflecting the complexity of the real world, the various models of bundling and tying differ greatly in their structure and assumptions. Moreover, these models generate nuanced and varied predictions.[10] While more research on these practices

[10] See, for example, Whinston (1990), Church and Gandal (2000), Denicolò (2000), and Nalebuff (2004). Rey and Tirole (2007) survey the literature on vertical foreclosure. For a recent analysis, see Chen and Rey (2023).

would be useful to help guide antitrust policy, robust and general results do not seem to be available.

Practical Approaches to Evaluating Cross-Market Acquisitions

Given these complexities, how should the antitrust agencies handle a cross-market acquisition that may allow the merged firm to gain a sizeable share in a new market? Four possible approaches come to mind.

First, the antitrust agencies could announce that they will challenge a cross-market acquisition only in cases where they can show that the bundling or tying enabled by the acquisition is likely to take place and to harm customers due to diminished competition. The 2023 Merger Guidelines do not appear to adopt this approach, although it is hard to tell based on the text. The European Commission, following this approach since 2008, has challenged very few cross-market acquisitions. The key policy question is whether that result reflects accurate enforcement or underenforcement. I am not aware of evidence showing significant harm to customers from cross-market mergers in Europe, but such harms would be difficult to detect even if they had arisen.

Second, the antitrust agencies could adopt a wait-and-see approach by letting an acquisition proceed if they are uncertain about whether the anticipated bundling or tying will occur and, if so, whether it will harm customers. Later, they could bring an enforcement action against the merged firm if it actually engaged in anticompetitive conduct that harmed customers, either to stop that practice or unwind the acquisition. Under this approach, the agencies would not have to predict post-merger conduct and its effects on competition and customers; instead, they would use post-merger evidence to challenge bundling or tying that harmed competition and injured customers. Perhaps the main drawback of this approach is that detecting potentially subtle changes in business practices and isolating the effects of these practices from other changes taking place over time can be very difficult. Thus, with a wait-and-see approach, antitrust agencies might not be able to detect and challenge harmful post-merger conduct even when it occurs.[11]

Under a third approach, the antitrust agencies could allow the cross-market merger to proceed subject to a commitment by the merged firm not to engage in certain bundling, tying, or other related conduct. This practice is known as a "behavioral remedy." For decades, the antitrust agencies have not accepted behavioral remedies for *horizontal* mergers. Instead, as a condition for approving certain horizontal mergers, they have insisted on divestitures, preferably of an entire line of

[11] The contrast here with horizontal mergers is instructive. Following a horizontal merger, a post-merger price increase by the merged firm is not itself prohibited under antitrust law unless done in concert with rivals. Moreover, illegal post-merger collusion might be very difficult to detect. For these reasons, the antitrust agencies do not permit horizontal mergers to proceed and plan to sue later if the merged firm raises price.

business rather than piecemeal assets. That policy is warranted because behavioral remedies—which might better be called ongoing regulatory rules—cannot deliver the broad and often unpredictable benefits of competition. By way of contrast, the antitrust agencies often *have* accepted behavioral remedies to resolve concerns for vertical mergers. However, the current leaders of the antitrust agencies have expressed deep skepticism about "fixing" mergers using remedies of any type, especially behavioral remedies (as opposed to "structural remedies" that involve asset divestitures).[12] The extent to which behavioral remedies for cross-market acquisitions can be effective in protecting and preserving competition is an open question.

A fourth approach to cross-market acquisitions would be for the antitrust agencies to just say "no" to many such proposed deals. Under this aggressive approach, the antitrust authorities would challenge cross-market acquisitions that risked extending a dominant position from one market into another, even without any showing of harm to customers in the new market. The agencies also would not accept behavioral remedies. This approach seems unwarranted, given the reasons noted above to expect that many cross-market acquisitions will benefit consumers and spur competition.[13]

Conclusion

With the 2023 Merger Guidelines, the US Department of Justice and Federal Trade Commission have announced an intention to go beyond horizontal and vertical mergers by scrutinizing cross-market acquisitions more closely. That expansion vastly increases the number of transactions that might be challenged by the antitrust agencies. We know that cross-market acquisitions have historically played a major role in contributing to the dynamism of the American economy by helping to create new markets and by disrupting powerful incumbents in established markets. Yet the 2023 Merger Guidelines say surprisingly little about such pro-competitive effects or how the antitrust agencies will actually analyze cross-market acquisitions.

The risk of counterproductive outcomes can be reduced if the antitrust agencies supplement the 2023 Merger Guidelines by acknowledging the benefits associated

[12] For example, Jonathan Kanter (2022), the leader of the Antitrust Division at the US Department of Justice, has stated: "I am concerned that merger remedies short of blocking a transaction too often miss the mark. Complex settlements, whether behavioral or structural, suffer from significant deficiencies. Therefore, in my view, when the division concludes that a merger is likely to lessen competition, in most situations we should seek a simple injunction to block the transaction. It is the surest way to preserve competition." Moss (2024) shows that the antitrust agencies have entered into far fewer settlements to resolve merger challenges during the Biden Administration than in previous administrations.

[13] Legislation recently introduced in the US Senate, the Competition and Law Enforcement Reform Act of 2024 (S.4308) would mandate an aggressive approach to cross-market acquisitions. That proposed law would stop large firms from making acquisitions by establishing a presumption that any acquisition of at least $50 million by a firm with assets, net annual sales, or market capitalization of at least $100 billion is illegal.

with cross-market acquisitions and by bringing the insights of economic research to bear to explain more fully how cross-market acquisitions will be analyzed.

References

Amgen. 2023. "Amgen Responds to FTC Action Re: Proposed Acquisition of Horizon Therapeutics." May 16. https://www.amgen.com/newsroom/press-releases/2023/05/amgen-responds-to-ftc-action-re-proposed-acquisition-of-horizon-therapeutics.

Areeda, Phillip, and Donald Turner. 1978. *Antitrust Law: An Analysis of Antitrust Principles and Their Application.* Boston: Little, Brown and Company.

Arnold, Daniel, Jaime S. King, Brent D. Fulton, Alexandra D. Montague, Katherine L. Gudiksen, Thomas L. Greaney, and Richard M. Scheffler. 2024. "New Evidence on the Impact of Cross-Market Hospital Mergers on Commercial Prices and Measures of Quality." *Health Services Research.* https://doi.org/10.1111/1475-6773.14291.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. 2014. "Vertical Integration and Input Flows." *American Economic Review* 104 (4): 1120–48.

Bauer, Joseph P. 1978. "Challenging Conglomerate Mergers under Section 7 of the Clayton Act: Today's Law and Tomorrow's Legislation." *Boston University Law Review* 58 (2): 199–245.

Bauer, Joseph P. 1983. "Government Enforcement Policy of Section 7 of the Clayton Act: Carte Blanch for Conglomerate Mergers?" *California Law Review* 71: 348–75.

Bet, Germán. 2024. "The Productivity Effects of Corporate Diversification." *Journal of Industrial Economics* 72 (2): 685–728.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122 (4): 1351–1408.

Bloom, Nicholas, and John Van Reenen. 2010. "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–24.

Chandler, Alfred D., Jr. 1977. *The Visible Hand: The Managerial Revolution in American Business.* Cambridge, MA: Belknap Press, Harvard University Press.

Chandler, Alfred D., Jr. 1990. *Scale and Scope: The Dynamics of Industrial Capitalism.* Cambridge, MA: Belknap Press, Harvard University Press.

Chandler, Alfred D. 1992. "Organizational Capabilities and the Economic History of the Industrial Enterprise." *Journal of Economic Perspectives* 6 (3): 79–100.

Chen, Zhijun, and Patrick Rey. 2023. "A Theory of Conglomerate Mergers." Toulouse School of Economics Working Paper 1447.

Church, Jeffrey. 2008. "Conglomerate Mergers." *Issues in Competition Law and Economics* 2: 1503–52.

Church, Jeffrey, and Neil Gandal. 2000. "Systems Competition, Vertical Merger, and Foreclosure." *Journal of Economics and Management Strategy* 9 (1): 25–51.

Dafny, Leemore, Kate Ho, and Robin Lee. 2019. "The Price Effects of Cross-Market Mergers: Theory and Evidence from the Hospital Industry." *Rand Journal of Economics* 50 (2): 286–325.

De Locker, Jan and Chad Syverson. 2021. "An Industrial Organization Perspective on Productivity." In *Handbook of Industrial Organization*, Vol. 4, edited by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, 141–223. Amsterdam: North-Holland.

Denicolò, Vincenzo. 2000. "Compatibility and Bundling with Generalist and Specialist Firms." *Journal of Industrial Economics* 48 (2): 177–88.

European Commission. 2001. "Case No COMP/M.2220—General Electric/Honeywell." July 3. https://ec.europa.eu/competition/mergers/cases/decisions/m2220_en.pdf.

European Commission. 2008. "Guidelines on the Assessment of Non-horizontal Mergers under the Council Regulation on the Control of Concentrations between Undertakings (2008/C265/07), Journal of the European Union." October 18, 265/6-25.

European Commission. 2018. "Case M.8394—Essilor/Luxottica." March 1. https://ec.europa.eu/competition/mergers/cases½02044/m8394_4245_3.pdf.

European Court of First Instance. 2005. "General Electric vs. Commission of the European Communities," CASE T-21%1, December 14, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62001TJ0210.

Federal Trade Commission. 2023. "In the Matter of Amgen Inc. and Horizon Therapeutics plc." Public Decision and Order Docket 9414. December 13. https://www.ftc.gov/system/files/ftc_gov/pdf/d09414amgenhorizonfinalorderpublic.pdf.

Jin, Ginger Zhe, Mario Leccese, and Liad Wagman. 2023. "M&A and Technological Expansion." *Journal of Economics and Management Strategy* 33 (2): 338–59.

Kanter, Jonathan. 2022. "Assistant Attorney General Jonathan Kanter of the Antitrust Division Delivers Remarks to the New York State Bar Association Antitrust Section." Speech, Washington, DC, January 24, 2022. https://www.justice.gov/opa/speech/assistant-attorney-general-jonathan-kanter-antitrust-division-delivers-remarks-new-york.

Kolasky, William J. 2002. "Conglomerate Mergers and Range Effects: It's a Long Way from Chicago to Brussels." *George Mason Law Review* 10 (3): 533–50.

Lee, Gwendolyn K., and Marvin B. Lieberman. 2010. "Acquisition vs. Internal Development as Modes of Market Entry." *Strategic Management Journal* 31 (2): 140–58.

Lee, Gwendolyn K., and Marvin B. Lieberman. 2024. "Exploration, Exploitation, and Mode of Market Entry: Acquisition versus Internal Development by Amazon and Alphabet." *Industrial and Corporate Change* 33 (1): 253–67.

Lewis, Matthew S., and Kevin E. Pflum. 2017. "Hospital Systems and Bargaining Power: Evidence from Out-of-Market Acquisitions." *Rand Journal of Economics* 48 (3): 579–610.

Moss, Diana L. 2024. "Taking Stock of Merger Enforcement under the Biden Administration," Progressive Policy Institute. https://www.progressivepolicy.org/wp-content/uploads/2024/09/PPI_Merger-Enforcement_V3.pdf.

Nalebuff, Barry. 2004. "Bundling as an Entry Barrier." *Quarterly Journal of Economics* 119 (1): 159–87.

Neal, Phil C., William F. Baxter, Robert H. Bork, and Carl H. Fulda. 1968. "Report of the White House Task Force on Antitrust Policy." *Antitrust Law and Economics Review* 11.

Nocke, Volker, and Michael D. Whinston. 2022. "Concentration Thresholds for Horizontal Mergers." *American Economic Review* 112 (6): 1915–48.

Patterson, Donna E., and Carl Shapiro. 2001. "Transatlantic Divergence in *GE/Honeywell*: Causes and Lessons." *Antitrust* 16 (1): 18–26.

Peters, Craig T. 2014. "Bargaining Power and the Effects of Joint Negotiation: The 'Recapture Effect.'" Economic Analysis Group Discussion Paper 14-3.

Ravenscraft, David J., and F. M. Scherer. 1987. *Mergers, Sell-Offs, and Economic Efficiency*. Washington, DC: Brookings Institution.

Ravenscraft, David J., and F. M. Scherer. 1989. "The Profitability of Mergers." *International Journal of Industrial Organization* 7 (1): 101–16.

Rey, Patrick, and Jean Tirole. 2007. "A Primer on Foreclosure." In *Handbook of Industrial Organization*, Vol. 3, edited by Mark Armstrong and Robert Porter, 2145–2220. Amsterdam: North-Holland.

Rose, Nancy L., and Carl Shapiro. 2022. "What Next for the Horizontal Merger Guidelines." *Antitrust* 36 (2): 4–13.

Schoar, Antoinette. 2002. "Effects of Corporate Diversification on Productivity." *Journal of Finance* 57 (6): 2379–2403.

Shapiro, Carl. 2001. "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting." *Innovation Policy and the Economy* 1: 118–50.

Shapiro, Carl. 2021. "Vertical Mergers and Input Foreclosure: Lessons from the *AT&T/Time Warner* Case." *Review of Industrial Organization* 59 (2): 303–41.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49 (2): 326–65.

United States. 2020. "Conglomerate Effects of Mergers." Note by the United States to the Competition Committee, Directorate for Financial and Enterprise Affairs, Organisation for Economic Cooperation and Development. https://www.ftc.gov/system/files/attachments/us-submissions-oecd-2010-present-other-international-competition-fora/oecd-conglomerate_mergers_us_submission.pdf.

United States Department of Justice, and Federal Trade Commission. 2020. "Vertical Merger Guidelines." https://www.justice.gov/atr/page/file/1290686/dl?inline.

United States Department of Justice, and Federal Trade Commission. 2023. “Merger Guidelines.” https://www.justice.gov/d9/2023-12/2023%20Merger%20Guidelines.pdf.

Vistnes, Gregory S. 2024. “Cross-Market Hospital Mergers: Assessing Likely Harm and Implications for Government Action.” *Antitrust Law Journal* 86 (1): 251–315.

Whinston, Michael D. 1990. “Tying, Foreclosure, and Exclusion.” *American Economic Review* 80 (4): 837–59.

Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* New York: Free Press.

Yip, George S. 1982. “Diversification Entry: Internal Development versus Acquisition.” *Strategic Management Journal* 3 (4): 331–45.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 77–100

Two Histories of the Public Safety Net

Christopher Howard

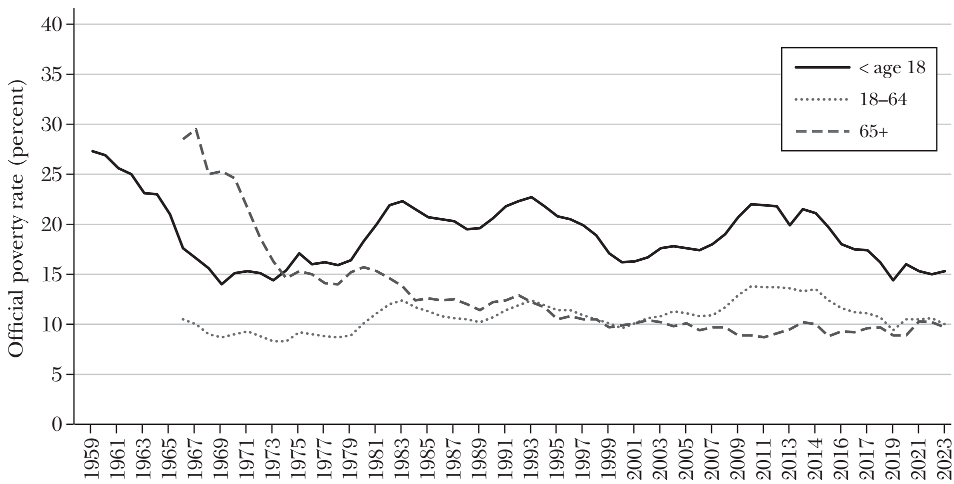
The poverty rate has been the leading indicator of material hardship for decades. In 1960, the official poverty rate was 22 percent, representing 40 million Americans. The poverty rate has been cut in half since then, a sign of substantial progress. Poverty has declined faster for some groups than others. The share of older Americans living in poverty dropped from 30 percent in the mid-1960s to 10 percent today, as shown in Figure 1. Poverty rates among adults aged 18–64 have changed much less. Child poverty rates have gone up and down, ultimately producing a larger decline than among working-age adults but smaller than among the elderly. Despite these improvements, 37 million Americans had incomes below the poverty line in 2023, and millions more lived near poverty (Shrider 2024). Compared to other affluent countries, the United States has had unusually high poverty rates for decades (Brady 2023).

Government programs that alleviate poverty are commonly referred to as the safety net. In this article, I will refer to these programs as the public safety net to distinguish them from the efforts of charities and extended families (Howard 2023). Though common now, the phrase "safety net" was not widely used in the United States until the 1980s (Safire 1981). Before then, policymakers were accustomed to discussing "poor relief" or "public assistance." The difference is important: for most of American history, officials never implied they would catch everyone who had fallen on hard times and protect them from hardship. The objective was to make life less miserable for some of the poor.

■ *Christopher Howard is the Pamela C. Harriman Professor of Government and Public Policy, College of William & Mary, Williamsburg, Virginia. His email address is cdhowa@wm.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20231393.

Figure 1
US Poverty Rates by Age



Source: Shrider (2024), Table A-3.

The public safety net is often equated with programs that are targeted at people with low incomes (Tach and Edin 2017; McKernan, Ratcliffe, and Braga 2021; Rachidi, Weidinger, and Winship 2022). In that case, the relevant history extends back to colonial times. Modern examples include Medicaid, Temporary Assistance for Needy Families, the Supplemental Nutrition Assistance Program (formerly Food Stamps), and public housing. The national government currently spends over $1 trillion on these programs each year (Center on Budget and Policy Priorities 2024). However, one can make a strong argument that inclusive social insurance programs also constitute a vital part of the public safety net (McGarry 2013; Moffitt 2013; Fishback 2018). Social Security, Medicare, and Unemployment Insurance are well-known examples, and they originated in the middle decades of the twentieth century. Though widely available, social insurance programs offer considerable support to individuals with lower incomes. Social Security currently lifts more Americans out of poverty than any other social program (Shrider 2024). Social Security is the single largest item in the federal budget, and annual spending on social insurance exceeds $2 trillion (Center on Budget and Policy Priorities 2024).[1]

By reputation, these two types of programs generate different levels of political support (Heclo 1986; Skocpol 1991; Hiltzik 2021). Means-tested programs are

[1] The public safety net relies on other policy tools as well. A longer article could have included regulations such as minimum wage laws and the Americans with Disabilities Act. With a few notable exceptions, tax expenditures generally do little to help people with low incomes. They subsidize pensions, health insurance, and housing for more affluent Americans (Howard 2007).

viewed as politically vulnerable, in part because targeting aid to a small share of the population limits the number of potential advocates. This problem is magnified when recipients lack income and political clout. State governments help run many of these programs, which creates obstacles to expansion and leads to cross-state variation. Some means-tested programs appear to undermine core values regarding work and family. Over the years, recipients have been stigmatized as freeloaders and welfare cheats. Negative racial, ethnic, and gender stereotypes have compounded these problems. For their part, inclusive social insurance programs are politically powerful. They benefit the rich, the poor, and everyone in between. Eligibility and benefits are tied to employment, reinforcing the work ethic and recipients' sense of entitlement. The stereotypical recipient has been white and male. As a result, elected officials have been more inclined to expand social insurance programs.

As we will see, a closer look reveals important similarities. The 1930s and 1960s were pivotal decades for means-tested and inclusive programs. Over the years, both have been criticized for doing too little to reduce poverty and for giving people more than they need. Some types of social insurance (like Unemployment Insurance) vary substantially from state to state, reinforcing the significance of federalism. At least one means-tested program, the Earned Income Tax Credit (EITC), is run exclusively by the national government.

This article charts the development of the public safety net, starting with means-tested programs. Over time, programs targeted at people with low incomes gradually shifted from the local to the state to the national level. Nevertheless, policymakers constantly made judgments about deservingness, and they often tried to limit cash welfare. Those judgments were less salient with inclusive social insurance programs, which expanded rapidly between 1950 and 1980, largely to the benefit of older Americans. The concluding section will highlight recent trends that challenge the classic distinctions between targeted and inclusive social programs. To help readers track these developments, Table 1 offers a chronology of important milestones.

A Brief History of Means-Tested Programs

The practice of poor relief in the United States dates to the seventeenth century. By mid-century, "so many indigent colonists were about—idlers, misfits, tramps, criminals, the sick, and so on—especially in growing towns like Boston, Newport, and Philadelphia, that their care had to become a community function and responsibility" (Trattner 1974, p. 16). Colonial assemblies passed laws, modeled after the 1601 English Poor Law, obliging localities to care for the destitute if family members could not. To limit the potential burden, local laws defined who qualified as a resident and how communities could deter undesirable individuals from settling in. Poor relief was for neighbors, not strangers. Those who were deemed eligible might receive indoor or outdoor relief, and the specifics varied from place to place. Indoor relief was provided inside someone else's residence. The poor could be temporarily housed with local families, who were compensated out of

Table 1
Milestones in the Public Safety Net

Time period	*Means-tested*	*Inclusive social insurance*
19th century	At the local level, indoor relief (like poorhouses) and outdoor relief (for example, food)	n/a
1910s–1920s	Mothers' pensions in most states; old-age pensions in some states	Workmen's compensation in most states
1930s	Social Security Act of 1935 created Aid to Dependent Children (ADC) and Old-Age Assistance; public housing began in 1937	Social Security Act of 1935 created Social Security (aka Old-Age Benefits, Old Age Insurance) and Unemployment Insurance
1940s–1950s	Considerable variations by state in assistance programs	Disability Insurance (1956); Large expansions in eligibility for Social Security
1960s–1970s	ADC became Aid to Families with Dependent Children (AFDC) and the program grew rapidly, prompting calls for welfare reform; Food Stamps (1964); Medicaid (1965); Supplemental Security Income (1972); Earned Income Tax Credit (1975)	Medicare (1965); Social Security benefits increased several times, then indexed for inflation; large decline in elder poverty
1980s–1990s	Cuts to AFDC (1981, 1996); gradual expansion of Medicaid and the EITC	Efforts to slow the growth of Medicare and Social Security
Early 21st century	Temporary expansion of assistance to cope with Great Recession and Covid-19 pandemic; Patient Protection and Affordable Care Act (2010) expanded Medicaid coverage; large decline in child poverty, starting in 1990s	Prescription drug benefit added to Medicare (2003); temporary expansion of Unemployment Insurance to cope with Great Recession and Covid-19 pandemic

public funds. Poor children, especially if they were orphans, could become indentured apprentices. In some places the poor were auctioned off, with the winning bidder accepting the smallest public subsidy. Outdoor relief was provided to individuals in their own dwellings. Food, clothing, or medical care might be donated by the community. The poor could be granted tax relief or allowed to plant gardens on public land. Direct cash assistance was rare; the poor could not be trusted to spend it wisely (Trattner 1974; Jensen 2015).

Cities created almshouses/poorhouses or workhouses in the seventeenth and eighteenth centuries, and many smaller towns and counties followed suit by the nineteenth century. This was indoor relief on a larger scale. In theory, poorhouses served people who were not expected to work, such as the elderly, disabled, and mothers with young children. Poor able-bodied adults were sent to workhouses and paid very low wages. In practice, many communities could afford to build just one of these institutions. Over time, workhouses were phased out and poorhouses remained, though they often compelled some residents to work. Living conditions were generally awful, as bad or worse than prisons. This was no accident; only the most desperate individuals would apply, and everyone else would get help from relatives or accept menial jobs. Otherwise, various forms of outdoor relief, seldom generous, were available from the seventeenth through the nineteenth centuries. Besides governments, churches and secular charities provided outdoor relief (Trattner 1974; Katz 1986).

The mix of indoor and outdoor relief varied across and within states. According to the 1880 US Census, there were three times more paupers in almshouses—tellingly referred to as “inmates”—than outdoor paupers. Almost all paupers resided in almshouses in states such as California and Ohio, whereas outdoor paupers were the majority in Texas, Vermont, and a few other states (US Census Office 1881). That same year, Kings County in New York (that is, Brooklyn) reported that all its paupers were in poorhouses; Erie County, including Buffalo, had less than one-third of its paupers in poorhouses. The mix varied over time as well. New York had more paupers receiving outdoor relief in 1850, 1860, and 1870, but more in poorhouses in 1840, 1880, and 1890 (Katz 1986).

When allocating relief, local officials did not treat everyone equally. The practice of driving away newcomers who might need assistance continued into the nineteenth century. In New England, unmarried mothers, Native Americans, and African Americans were more likely to be excluded (Herndon 2001). Likewise, poor immigrants from countries such as Ireland and Germany were scrutinized closely when they sought aid. In the South, where most African Americans lived, slaves were the responsibility of their owners, not the wider community, and free Blacks were usually ineligible for local relief. No matter where they lived, able-bodied men were commonly viewed as lazy or immoral if they requested public assistance (Trattner 1974; Katz 1986; Jensen 2015).

Both indoor and outdoor relief were widely criticized by the end of the nineteenth century. Poorhouses were inhumane and expensive. Outdoor relief fostered dependence, some claimed; outdoor relief was too meager, said others. Neither type of relief was staffed properly, leading to mismanagement and inefficiencies. Neither approach addressed the underlying causes of poverty. The main alternative was organized charity, but the gap between social needs and available resources was enormous (Lowell 1884; Trattner 1974; Katz 1986).

New Forms of Public Assistance

In the first decades of the twentieth century, the number of poorhouses declined while outdoor relief gradually expanded. Cities and counties were still involved, local officials still distinguished between the deserving and undeserving poor, and the amount of aid was still minimal. Blacks seldom received public assistance in the South. Mexicans and Mexican Americans were increasingly criticized for being lazy and dependent in the West and Southwest. Immigrants from southern and eastern parts of Europe found it easier to receive poor relief in the Northeast and Midwest, partly because social workers were sympathetic to their plight (Fox 2012). Simultaneously, state governments started to play a larger role and cash assistance became more common. The main beneficiaries were poor single-mother families and the elderly poor.

Mothers' pensions were the earliest form of "welfare," which later became the most controversial part of the public safety net. Most states enacted these pensions during the 1910s. Among the key advocates were women's associations, especially the General Federation of Women's Clubs and the National Congress of Mothers, and juvenile court judges who often dealt with poor families. Cash assistance would keep poor single-mother families together, reducing the number of children sent to orphanages and foster care. The chief opponents of mothers' pensions were organized charities who felt that government aid too often fostered dependence. The poor, in their view, required careful supervision to become self-sufficient. While states established general rules for eligibility, they allowed county governments to decide whether to offer mothers' pensions and how much each pension would be worth. Counties were also expected to finance mothers' pensions without state funds (Howard 1992; Skocpol 1992). Local responsibility, a hallmark of nineteenth-century poor relief, remained a guiding principle.

Geographic differences were substantial. Almost all states covered single mothers who were widowed, roughly half the states covered mothers whose husbands had divorced them, and about one-quarter covered mothers who had never married. State laws also allowed local officials to determine the fitness of the mother and the suitability of her home, which were highly subjective. A mother's housekeeping skills, use of alcohol and tobacco, and sexual behavior could be investigated. In the early 1920s, the US Children's Bureau found that the program was biased in favor of Anglo-Saxon mothers, whose behavior came closer to what caseworkers thought proper. A national survey in 1931 found that Black families were badly underrepresented among recipients. Some Southern states had no Black recipients of mothers' pensions, and others had just a handful (Bell 1965; Howard 1992; Gordon 1994).

That same survey revealed large variations in benefits. The median monthly pension was \$20–\$25 per family. The average was \$10 per family in Florida, Louisiana, and Texas but much higher in New York and close to \$70 per family in Massachusetts. Like poor relief in the nineteenth century, the amounts varied within states. In Illinois, the average monthly benefit was \$21 in 1933, but that included many counties where benefits were less than \$10 and a few where they exceeded \$30. Differences in cost of living explain some of these differences, but not all. Local officials could choose whether to give smaller benefits to more families or larger benefits to fewer families. Officials often favored "respectable" mothers, thus allowing local prejudices to influence benefit levels (Howard 1992; Skocpol 1992).

Most poor single mothers never benefitted from mothers' pensions, either because their county did not offer them or because they were stuck on a waiting list for scarce funds. Barely 3 percent of all single-mother families received mothers' pensions in 1931 (Patterson 1994). Even single mothers living in the most generous states struggled to make ends meet. Many mothers supplemented their pensions with part-time jobs because state laws usually required mothers to be home with their children at least four days per week. Some recipients took in laundry or sewed clothes at home. From the beginning, poor women had to combine welfare with low-wage labor (Howard 1992; Goodwin 1997; Ward 2005).

This outcome was somewhat ironic. Initially, advocates argued that mothers provided a valuable service to society by raising the next generation of workers and citizens. Mothers should be valued like soldiers and given pensions like veterans. Poor single mothers were not simply another needy group; they had earned their benefits (Orloff 1991; Skocpol 1992). Those arguments might have helped enact these laws, but they had little impact on implementation. Most of the groups that worked hard to create mothers' pensions moved on to fight other battles (Howard 1992; Skocpol 1992).

Between the early 1920s and early 1930s, many states enacted pensions for the elderly poor. Officials were dismayed to see poorhouses functioning as ramshackle old-age homes. Cash assistance would help the elderly live with a little more dignity. However, eligibility rules could be strict: recipients in Ohio had to prove residency of at least 15 years; show they had less than $300 of annual income; pass tests of moral fitness; and give the state title to any home they owned (Berkowitz 1991). Few Americans received old-age pensions before 1929, but by 1932 these pensions were comparable to mothers' pensions in size, scope, and variation from state to state. One notable difference was that several states required old-age pensions to be offered in every locality (Lubove 1968; Béland 2005).

A New Deal for the Poor

The enormous hardships generated by the Great Depression paved the way for the national government to become more involved in means-tested programs (Witte 1963; Mettler 1998). The three main programs available from the New Deal of the 1930s until the Great Society of the 1960s were Aid to Dependent Children, Old-Age Assistance, and public housing.

The Social Security Act of 1935 is best known for creating Social Security, which will be discussed later, but it also targeted assistance to some groups with low incomes. Aid to Dependent Children (ADC) benefitted poor single-mother families while Old-Age Assistance (OAA) assisted the low-income elderly. These parts of the Act were not that controversial because government's role in aiding these groups had been gaining acceptance, and because state governments would remain pivotal in running the new programs (Witte 1963). Moreover, the initial outlays were small. The national government was obligated to spend $50 million per year on Old-Age Assistance, equal to a little over $1 billion today. The budget for Aid to Dependent Children was half that size (Social Security Administration n.d.).

The Social Security Act became law at a time when Franklin Roosevelt was President and Democrats enjoyed huge majorities in the House and Senate. Republicans could do little to resist. However, the southern and northern wings of the Democratic party did not always agree, and conservative Southern Democrats held powerful positions in Congress. Compromises and concessions were routine. For instance, the Roosevelt administration wanted Aid to Dependent Children and Old-Age Assistance to "afford a reasonable subsistence compatible with decency and health." That standard would have led to higher benefits in most states compared to their prior pension programs. Southern Democrats objected, using arguments that were often implicitly or explicitly racist: basically, if public assistance was too generous, then it would become harder to pay low wages to Black workers. Single mothers and the elderly could be hired as domestic or agricultural workers, even if on a seasonal or part-time basis. Some Southerners feared that federal regulation of Old-Age Assistance benefits would lead to greater interference in race relations. The final version of the Social Security Act allowed states to set benefit levels wherever they liked, and the national government would reimburse a portion of their spending (one-third for ADC, one-half for OAA). Southern Democrats also managed to delete a provision that would have required state welfare agencies to abide by the same personnel guidelines for hiring and compensation used by the national government. State agencies would still be run by people who had been approved by the state's politicians (Lieberman 1998; Mettler 1998; Ward 2005).

Public housing projects originated with the Housing Act of 1937, in which the national government would help local housing authorities construct and operate apartment buildings, aimed primarily at low-income individuals and families. The Great Depression put a serious dent in housing construction. During his first term, President Roosevelt initiated several efforts (like the Public Works Administration) to construct more dwellings and boost employment in the building trades. These projects tended to reinforce or aggravate residential segregation by race, not only in the South but all over the country (Rothstein 2017). To mollify southern Democrats, the act imposed a low ceiling on the cost of each public housing unit. Cheap construction would affect those individuals who could not afford to pay market rents or be picky about quality (Radford 1996; von Hoffman 2012).

In their first decades, Aid to Dependent Children, Old-Age Assistance, and public housing were growing, but hardly amounted to a safety net. By 1963, four million Americans, including three million children, were receiving Aid to Families with Dependent Children (the name changed in 1962). That was less than one-fifth of all poor children nationwide. For the average AFDC family, annual benefits were worth roughly half the poverty threshold (US Bureau of the Census 1964; US Census Bureau 2023a, 2023b). With so little cash assistance, poor single mothers still had strong incentives to find employment or get married. Old-Age Assistance was likewise limited: there were two million recipients in 1963, which was less than half the total number of poor elderly Americans; and the average annual benefits were well below the poverty threshold (US Bureau of the Census 1964; US Census Bureau 2023a, 2023b). Finally, the United States had approximately 550,000 units of public housing in 1963, with another 40,000 units under construction. The total poverty population was over 36 million (US Bureau of the Census 1964; US Census Bureau 2023a).

Geographic variation remained the norm. Twenty-one states, especially but not exclusively in the South, adopted work requirements for ADC/AFDC parents by the early 1960s. Other states did not (Soule and Zylan 1997). In 1963, monthly AFDC benefits were less than $20 per person in most Southern states, but close to $50 in Minnesota and Wisconsin. Old-Age Assistance benefits typically ranged from $50 to $100 per month, depending on the state (US Bureau of the Census 1964).

Although it is hard to quantify how much states relied on subjective “fitness” or “suitability” criteria, considerable evidence suggests that Southern states in the 1940s and 1950s made it harder for Black families to qualify for ADC and harder to receive benefits equal to white families (Bell 1965; Patterson 1994; Ward 2005). Public housing continued to be racially segregated. Liberal Democrats in Washington accepted this practice to generate enough support in Congress to finance additional projects (Rothstein 2017).

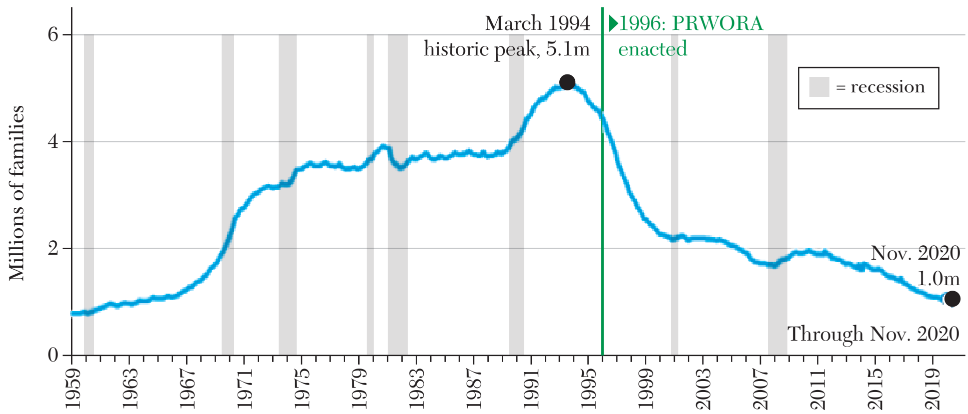
The War on Poverty and Its Aftermath

A few months after being elected president in 1964, Lyndon Johnson made headlines by calling for an unconditional “war on poverty.” The next few years were arguably more important than the 1930s in shaping means-tested programs. The enactment of Food Stamps (1964) and Medicaid (1965) was crucial. The Food Stamps program, which was initially optional for the states, took a while to spread across the country, but by the early 1990s it cost as much as Aid to Families with Dependent Children and served twice as many people (US Bureau of the Census 1994; Bosso 2023). Medicaid’s impact was more immediate. Millions of poor Americans, primarily those receiving AFDC and Old-Age Assistance, now had health insurance. They could obtain medical care more easily, and hospitals could be reimbursed for care they used to provide the poor for little or nothing (Olson 2010). By the late 1960s, total spending on Medicaid surpassed AFDC as well as OAA. Medicaid was a core part of the public safety net by 1980 (US Bureau of the Census 1972, 1982).

Neither Food Stamps nor Medicaid generated strong opposition at the outset. Similar to the New Deal era, Democrats in the mid-1960s controlled the White House and enjoyed large majorities in Congress, but the southern wing of the party still wielded power (Bertram 2015). Food Stamps moved forward as soon as advocates agreed to support agricultural subsidies, forging an urban-rural coalition that lasted for years (Bosso 2023). Medicaid was less controversial than Medicare, which became law at the same time. The American Medical Association feared that Medicare, not Medicaid, would lead to socialized medicine (Derthick 1979; Oberlander 2003). Smaller means-tested programs such as the Job Corps for young adults and Head Start for preschool children also originated in 1964–1965, without much pushback.

One means-tested program that President Johnson did not want to expand—Aid to Families with Dependent Children, aka “welfare”—grew rapidly and became a political lightning rod. Johnson thought that government should give the poor more goods and services than cash assistance. But welfare caseloads tripled between 1960 and the early 1970s, as shown in Figure 2, due to a combination of demographic shifts, court rulings, and organizations that helped the poor apply for and stay on welfare. This growth troubled officials from both political parties (Quadagno 1994; Davies 1996; Teles 1996). Calls for welfare reform became a prominent feature of American politics from the late 1960s until the end of the twentieth century.

Figure 2
Families Receiving AFDC/TANF, 1959–2020



Source: Congressional Research Service, based on data from the Department of Health and Human Services (Falk 2021).
Note: PRWORA refers to the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, which transformed AFDC into TANF.

Officials kept trying to distinguish between the deserving and undeserving poor. The poor elderly and poor disabled were considered worthy because they could not work enough to be self-sufficient. The Supplemental Security Income program, created in 1972, was designed to standardize and improve their benefits (Berkowitz and DeWitt 2017). Adult welfare recipients were seen as less deserving, and officials tried to move them into the labor force through work incentives, work requirements, or social services like job training and child care. One early reform package (in 1962) relied on a services approach. When that failed to stem the growth of Aid to Families with Dependent Children, national officials adopted tougher work policies in 1967 and 1972. (For obvious reasons, the new Work Incentive Program was called WIN, not WIP.) Starting in 1981, President Ronald Reagan gave states more flexibility to promote employment among AFDC parents. The Family Support Act of 1988, enacted with bipartisan support, added work requirements and services for welfare recipients. All these policy changes represented a shift from welfare to "workfare" (Weaver 2000; Bertram 2015).

Some of these changes produced significant cuts. President Reagan's first budget in 1981 reduced spending on a wide range of means-tested programs, including Aid to Families with Dependent Children. One common approach was to restrict eligibility by lowering income thresholds or tightening asset tests. At the state level, legislators in the 1970s and 1980s allowed inflation to erode the real value of AFDC benefits, which declined all over the country (Weaver 1988; Bertram 2015). These cuts had the effect of pushing more welfare recipients into the low-wage labor force.

During the twentieth century, cash welfare had seldom been popular. Officials at the state and national levels looked for ways to make welfare less appealing. The political climate became more hostile in the 1970s and 1980s as poverty-related issues became more racialized. Negative stories about poverty and welfare appeared more often in the media, disproportionately featuring pictures of Blacks (Gilens 1999). Public criticism of "welfare queens" and the underclass intensified. Welfare receipt was linked to a wide variety of social ills such as chronic joblessness, drug abuse, teen pregnancy, and out-of-wedlock births (Wilson 1987; Borelli 2019). Although the racial dimension was not always explicit (Mead 1992), some authors openly blamed the underclass on a Black culture of poverty (Kaus 1986).

In these disputes, the term "welfare" was ambiguous. Sometimes it referred only to Aid to Families with Dependent Children or to AFDC plus Food Stamps. At other times it signified all means-tested programs. Perhaps the most famous welfare critic of this era was Charles Murray (1984), author of the book *Losing Ground*. To Murray, "welfare spending" meant the combination of AFDC, Food Stamps, Medicaid, public housing, and similar programs. The growth of welfare spending since 1960, he argued, made it less rational for the poor to find a low-wage job. They could live off public assistance. The government was spending billions and billions of dollars on social programs that fostered welfare dependence instead of reducing poverty or promoting self-sufficiency. When President Reagan (1986) famously stated that America had declared war on poverty and poverty won, he was echoing Murray.

Amending Welfare as We Knew It

In the early 1990s, Bill Clinton was campaigning for the presidency and promising to "end welfare as we know it." Clinton was thinking mostly about changes to Aid to Families with Dependent Children, but Republicans had broader plans to reduce the public safety net. Leading up to the historic welfare reform act of 1996 (officially the Personal Responsibility and Work Opportunity Reconciliation Act), Republicans tried to convert food assistance programs into a block grant, giving states more leeway in running these programs while cutting federal spending. In 1994, the "GOP Contract with America" pledged to help states create more orphanages for poor children. Neither proposal became law. However, Republicans did succeed in excluding many legal immigrants from Food Stamps, Medicaid, and Supplemental Security Income in the 1996 welfare reform act. Although President Clinton criticized these exclusions, he had signaled a willingness to consider them in his 1994 welfare reform proposal (Weaver 2000; Singer 2004).

The 1996 welfare reform act replaced Aid to Families with Dependent Children with Temporary Assistance for Needy Families (TANF), and two changes stood out. First, funding was transformed from a budgetary entitlement, in which the federal government committed to making a share of payments to all eligible individuals, to a block grant to states. This change effectively capped federal outlays in the future (at $16.5 billion per year) and gave states more leeway in running the program. Second, adult welfare recipients now faced time limits, two years per spell and five years lifetime, though states could shorten those limits (and several did). Both changes were intended to move more of the poor from welfare to work. While the law also increased work requirements for TANF recipients, limited eligibility for legal immigrants, and tried to promote marriage and two-parent families, block grants and time limits had the biggest long-run impacts. Politically, the 1996 welfare reforms were notable because a Democratic president agreed to shrink the public safety net, much to the displeasure of liberal Democrats, and he worked with conservative Republicans to make that happen. Welfare reform was part of a larger effort by Clinton to move the Democratic Party in a more centrist direction (Weaver 2000; Bertram 2015).

Cash welfare has attracted much less attention in recent decades. Temporary Assistance for Needy Families caseloads dropped dramatically in the late 1990s, as shown in Figure 2, without increasing poverty rates, leading some to declare welfare reform a huge success (Haskins 2006; Kim and Rector 2006). The 1996 welfare reform act has been tweaked, but its basic shape remains. Neither the Great Recession of 2007–2009 nor the Covid-19 pandemic prompted officials to overhaul TANF. The persistence of strict time limits and tight eligibility criteria means that only a small fraction of poor families with children currently qualify. In absolute numbers, the welfare caseload in 2020 was comparable to 1960. Most states have chosen to reduce TANF cash assistance in favor of services that promote employment. Benefits still vary considerably around the country, and Black children are more likely to live in low-benefit states (Center on Budget and Policy Priorities 2022).

In short, what started out as a war on poverty morphed into a war on cash welfare. The evolution of Aid to Families with Dependent Children and Temporary Assistance for Needy Families illustrate the old saying, “programs for the poor are poor programs.” With so little support, these programs were vulnerable to attack.

Social Insurance and the Public Safety Net

Like other countries, the United States offered poor relief long before it established a welfare state. The advent of social insurance in Europe during the late nineteenth century marks the beginning of the modern welfare state. These new programs could help protect workers and their families against sudden drops in income that were common in industrial, capitalist economies. Social insurance would lessen the impact of unemployment and retirement, for example. Recipients might be poor, and officials often touted the ability of social insurance programs to reduce poverty, but there was no means test. Benefits would be earned through employment (Garland 2016; Howard 2023).[2]

When social insurance programs emerged in this country, they were layered on top of means-tested programs. Social Security and Unemployment Insurance

[2] Though the US government funded Civil War pensions in the late nineteenth and early twentieth centuries, these outlays were tied to military service rather than wage work (Skocpol 1992).

originated with the Social Security Act of 1935.[3] Both programs generated more controversy than Old-Age Assistance or Aid to Dependent Children. Poor relief had a long history, but social insurance did not. Relying on payroll taxes to fund government programs was novel, and the idea of imposing a new tax in the middle of the Great Depression struck many as foolish. To make matters worse, Social Security (originally called old-age benefits) would not send benefit checks until 1942, even though millions of older Americans needed help immediately. President Roosevelt, however, felt that the combination of payroll taxes now and benefits later would help instill a feeling that recipients had earned government's help. Social Security attracted criticism as the only truly national part of the Social Security Act, and Unemployment Insurance was conspicuous for giving money to working-age adults. Business and labor groups were much more interested in shaping Social Security and Unemployment Insurance than the means-tested programs in the Social Security Act (Mettler 1998; Kennedy 1999; Béland 2005).

Social Security, Medicare, and the Poor Elderly

Initially, Social Security had little impact on poverty. The delay in paying benefits was one reason, as was the plan to keep benefits small and expand them slowly. In addition, many occupations were originally excluded from Social Security and Unemployment Insurance. While most factory workers were included, agricultural workers and domestic workers were not, which affected many low-wage workers and most Black workers. The close link between employment and eligibility effectively excluded most women from both programs. Between 1936 and 1950, Social Security benefited fewer people and paid smaller benefits than Old-Age Assistance. Most Republicans and some Democrats in Congress believed that means-tested programs were the better way to help the elderly (Derthick 1979; Lieberman 1998; Mettler 1998).

The political virtues of inclusive social insurance took a few decades to materialize. Social Security did not qualify as the "third rail" of American politics until the 1970s and 1980s (Campbell 2003). Over the years, program benefits gradually expanded because of a small, dedicated group of bureaucrats and legislators, sometimes aided by organized labor (Derthick 1979; Zelizer 1998).[4] The first step was to broaden eligibility. Spouses, survivors, and dependents gained coverage in 1939, which meant that some of the poor widows with children who would have received Aid to Dependent Children—and who would have enhanced that program's public image—shifted over to Social Security (Mettler 1998). Most of the occupational

[3] Technically, workmen's compensation became the first form of social insurance in the United States when most states passed compensation laws in the 1910s and 1920s. However, this program remains firmly ensconced at the state level, the national government collects limited data about the program, and its antipoverty effects are poorly understood (Howard 2002).

[4] Some of these advocates worked hard to promote Social Security by contrasting it with Old-Age Assistance. The former was earned, they argued, while the latter was more like a handout. Programs for the poor may be poor programs partly because their reputation was tarnished by proponents of social insurance (Cates 1983; Zollars and Skocpol 1994).

exclusions were dropped during the 1950s, giving Social Security a cross-class constituency. The program was not and still is not truly universal, but by 1960 it was broadly inclusive. The addition of disability benefits in 1956, originally paid to workers between the ages of 50 and 65, helped those who were on track to collect Social Security but could no longer work (Derthick 1979).

The next steps involved expansion of Social Security benefits. In 1967, President Johnson called for higher benefits to help reduce poverty among the elderly, and Congress signed off on a 13 percent increase the following year. President Nixon and Congress continued down this path in 1969 (15 percent increase), 1971 (10 percent), 1972 (20 percent), and 1973 (7 percent). These numbers were well above inflation at the time. Starting in 1975, Social Security benefits were indexed to inflation, thereby locking in these gains (Derthick 1979; Béland 2005; McGarry 2013). In constant (2022) dollars, the average monthly benefit for a retired worker increased from $732 to $1,213 between 1960 and 1980 (Social Security Administration 2000). As Social Security grew, the elderly had less need for the means-tested Supplemental Security Income program, which increasingly served the disabled poor.

In addition, the benefit formula was designed so that Social Security would replace a larger share of pre-retirement wages for lower-wage workers, which gave them a better chance of avoiding poverty after they retired (Derthick 1979). This feature of Social Security is a classic example of "targeting within universalism" (Skocpol 1991), and it is vital given how little savings or other pension income these workers have when they retire (Li and Davies 2022).

As more Americans received bigger Social Security benefits, the program gained a devoted constituency. The American Association of Retired Persons (now called AARP) gained a reputation as one of the country's most powerful interest groups. When President Reagan proposed a sizable cut to Social Security benefits in 1981, the backlash was swift and loud, making elected officials very reluctant to cut benefits in the future. Instead, Reagan appointed a National Commission on Social Security Reform, chaired by Alan Greenspan. The Commission's report in 1983, along with some behind-the-scenes dealmaking, became the basis for the Social Security Amendments of 1983. These changes included a gradual increase in the normal retirement age from 65 to 67. For the first time, Social Security benefits would be subject to income taxation, but only for more affluent retirees. Otherwise, the main features of the program remained largely intact (Campbell 2003; Béland 2005; Howard 2007).

The addition of Medicare in 1965, financed partly by payroll taxes, gave millions of senior citizens a substantial boost. Previously, more than half of them lacked private health insurance, yet their medical expenses were twice as high as everyone else's. Many of the same advocates who pushed for Social Security's expansion also pushed for Medicare. While they favored a social insurance approach, they also believed that many of the uninsured elderly were poor. Unlike Social Security, the benefits of Medicare were widely available from the beginning. Anyone age 65 and over who qualified for Social Security would also be covered for hospital care (Medicare Part A), and they could choose to obtain more routine medical care under Part B of the program. Unlike Social Security, the size of Medicare benefits would not be linked to past wages; older Americans with low incomes could receive substantial amounts of medical care.[5] Fiscal conservatives in the early 1960s worried that Medicare's design would lead to uncontrollable growth (Derthick 1979; Oberlander 2003). Politically, the advent of Medicare reinforced the expansion of Social Security, giving senior citizens more motivation and resources to monitor these programs and influence policy change (Campbell 2003).

Not only did Social Security and Medicare become two of the largest programs in the national budget. They also gained much broader political support than cash welfare ever did. A Republican, President George W. Bush, led the push to add prescription drug benefits to Medicare in 2003. A Republican, Donald Trump, campaigned for president in 2016 and 2024 with promises not to cut a penny from Social Security or Medicare.

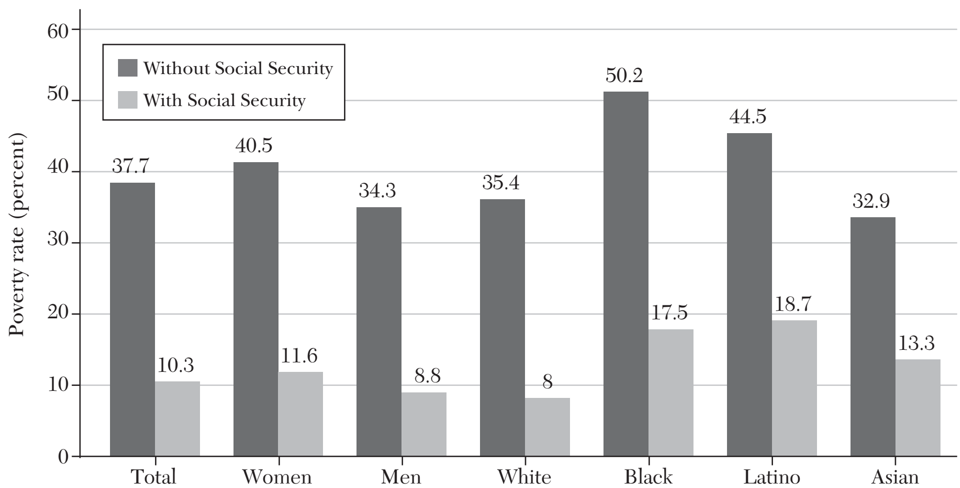
For elderly Americans, one big payoff was a historic decline in poverty. The official poverty rate for older Americans (age 65+) dropped by half in the 1960s and 1970s (US Census Bureau 2023a) at a time when their Social Security checks were growing, and Medicare had started paying their hospital bills (Campbell 2003; McGarry 2013). The elderly poverty rate continued to drop, albeit more slowly, until it reached 10 percent in 2000. During the twenty-first century, the elderly have been less likely to live in poverty than the nonelderly (as shown earlier in Figure 1).

Figure 3 shows how the impact of Social Security on the elderly cuts across gender, racial, and ethnic lines. These numbers do not reflect behavioral changes that might occur in the absence of Social Security, such as more saving for retirement or greater willingness to live with adult children. Without Social Security, poverty rates for older women and older men would have more than tripled in 2021. Social Security reduced the poverty rate for older Blacks from 50 to 18 percent, and for older Latinos from 45 to 19 percent. If we think of Social Security as Old-Age, Survivors, and Disability Insurance (OASDI), then the antipoverty effects extend to younger groups. Social Security lifted more than five million adults aged 18–64 above the poverty line in 2021 and did the same for one million children (Romig 2023).[6] In short, Social Security is an essential part of the social safety net, and its history and politics stand in stark contrast to cash welfare programs.

[5] In 1965–1966, a coalition of public officials and civil rights groups used Medicare and the 1964 Civil Rights Act to desegregate Southern hospitals. If hospitals wanted to be reimbursed by Medicare, they could not be segregated. Racial disparities in health outcomes (like infant mortality) diminished, especially in the South (Smith 2015). This unintended consequence of Medicare positively affected African Americans, many of them low income.

[6] For technical reasons, the Census Bureau has been unable to estimate how many people are lifted out of poverty by Medicare or Medicaid. This may change in the not-too-distant future (Dalaker 2022).

Figure 3
Poverty Rates for Age 65+ with and without Social Security (2021)



Source: Center on Budget and Policy Priorities, using data from US Census Bureau's March 2022 Current Population Survey (Romig 2023).
Note: Poverty rates are based on the Official Poverty Measure.

Unemployment Insurance and Poverty

The political dynamics of Unemployment Insurance (UI) have not followed the same path as Social Security or Medicare. One reason is that while most Americans expect to grow old, they do not necessarily expect to be unemployed—or at least not for long periods of time. Mobilizing people around a temporary misfortune is difficult. The program's design also makes it far less inclusive. Eligibility restrictions keep most of the unemployed from collecting UI benefits (for example, workers who quit voluntarily or held part-time jobs are not eligible). Benefits are paid out for months, not years. As a result, no mass membership group has emerged to represent the unemployed or recipients of UI.

The original Social Security Act gave states substantial authority to set eligibility criteria and benefit levels for Unemployment Insurance, and state-level variation has remained the norm. Agricultural and domestic workers were generally excluded from 1935 up to the 1960s, giving unemployed workers less help in rural states and making it harder for Black workers to qualify. In Texas, the unemployment insurance program did not include those occupations until 1985 (Lieberman 1998). In the mid-1990s, UI benefits would have replaced two-thirds of prior wages for unemployed workers in Oregon, but only two-fifths of prior wages in Nebraska (Graetz and Mashaw 1999). Shortly before the Covid-19 pandemic, basic UI benefits were available for up to 26 weeks in most states; up to 20 weeks in Michigan, Missouri, and South Carolina; and for 12 or 14 weeks in Alabama, Florida, Georgia, and North Carolina, depending on the state's unemployment rate (Whittaker and

Isaacs 2019). Over half of the unemployed in Minnesota and New Jersey received UI in 2022, compared with less than 10 percent in Alabama and North Carolina (Gwyn and Gerry 2023). Such variation is more characteristic of Temporary Assistance for Needy Families than Social Security. States have little incentive to make UI comprehensive or generous. Benefits typically replace less than half of workers' earnings. A limited safety net will push unemployed workers to take whatever jobs are available (Graetz and Mashaw 1999; Stone and Chen 2014; Woodbury 2015).

When the economy is reasonably healthy, Unemployment Insurance has little impact on poverty. Unemployment benefits lifted only 300,000 people out of poverty in 2023, which was much less than the means-tested SNAP program (Shrider 2024). UI is more important during economic downturns, but that is due primarily to the extended or emergency benefits that are paid by the federal government. These additional benefits were essential when the Great Recession and the pandemic overwhelmed states' UI programs (Howard 2023).

Discussion

Given the stark contrasts between means-tested "welfare" and inclusive Social Security, it is not surprising when advocates for paid family and medical leave (Romig and Bryant 2021) or long-term care (Powell 2019) endorse a social insurance model. Nor is it surprising when they endorse Medicare for All (Scott 2019). Part of me is attracted to these proposals, for reasons that range from lower administrative costs to social solidarity. I am also attracted by the possibilities of giving extra help to lower-income households within broadly inclusive programs. However, recent history does not point decisively in favor of inclusive over means-tested programs.

In retrospect, the biggest achievements of social insurance, politically and programmatically, occurred between the 1930s and 1970s. Since then, Social Security and Medicare have had more success preventing retrenchment than accomplishing expansion. Recent debates over "saving" Social Security typically focus on raising revenues or cutting the trajectory of future spending, not adding benefits (Arnold 2022). Similar debates over Medicare usually focus on cost control. For decades, national officials have been reluctant to increase payroll taxes for new or existing social insurance programs. Republicans do not want to raise taxes, period, and many Democrats worry about the regressivity of payroll taxes (Campbell and Morgan 2005). For example, payroll taxes would not have financed the Clinton health plan in the early 1990s or Biden's Build Back Better initiative in the early 2020s, and they did not fund Medicare's prescription drug benefit that was added in 2003. The Patient Protection and Affordable Care Act of 2010 did raise payroll taxes for Medicare, but only on the most affluent workers. In short, broadly inclusive programs can lead to spending levels that make further expansion difficult.

Now consider the history of means-tested programs besides AFDC/TANF during the last half-century. Since the 1970s, several of these programs became more uniform across the country, expanded, or both. Changes to Food Stamps in the

1970s made eligibility standards and benefits the same in every state. Eligibility and benefits gradually increased between 1985 and 1990, and the program expanded again as part of two farm bills during the George W. Bush administration in the early 2000s (Greenstein 2022; Bosso 2023). The introduction of Supplemental Security Income (SSI) reduced variation in eligibility and benefits compared to previous programs for the poor elderly (as well as for the low-income blind and disabled) under Old-Age Assistance. By 1995, the national government was spending just as much on SSI as on Food Stamps, and both were larger than AFDC. Between 1974 and 2022, SSI benefits grew from $5 billion to $56 billion; adjusted for inflation, spending almost doubled (US Bureau of the Census 2000; Social Security Administration 2023).

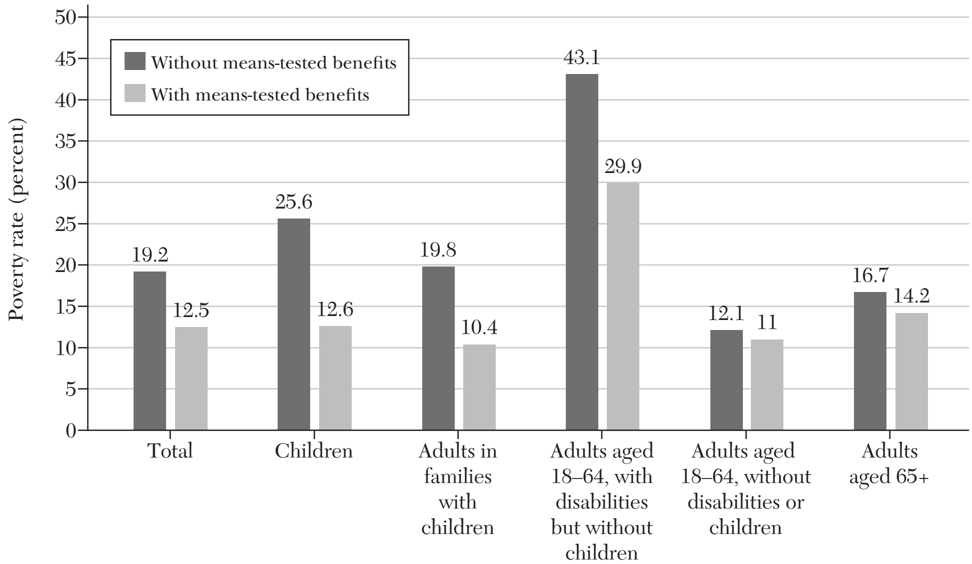
The Earned Income Tax Credit (EITC) originated in 1975 and grew even faster. The EITC is targeted at wage earners with low incomes, roughly up to twice the poverty line. It reduces their income tax liability and often generates a tax refund. Due to expansions in 1986, 1990, and 1993, spending on the EITC surpassed AFDC/TANF by the late 1990s. In 2022, the EITC was claimed on 27 million tax returns (most of which represented families of two or more individuals) and cost the US government $65 billion. As a program run through the federal income tax, it is administered solely by the national government (Howard 2007; Joint Committee on Taxation 2022).

Officials expanded means-tested medical care as well. A series of changes enacted between 1984 and 1990 extended Medicaid coverage to millions of pregnant women, new mothers, and children who were poor but had not previously been ineligible for Aid to Families with Dependent Children. Like the Earned Income Tax Credit expansions mentioned above, Medicaid expansions were usually tucked away in massive tax or budget bills where they attracted little attention. Advocates for the poor were involved, but not the poor themselves. The Children's Health Insurance Program (CHIP), enacted in 1997 with bipartisan support, extended coverage to low-income children whose families were ineligible for Medicaid (Howard 2007).

The Patient Protection and Affordable Care Act of 2010 expanded means-tested Medicaid, not social insurance Medicare, to cover more of the uninsured (Starr 2013). However, states had to choose to participate, and because several of the holdouts had large populations of low-income Blacks and Hispanics (including Georgia, Florida, and Texas), the law did less to close racial gaps in health insurance than hoped. Even so, almost 75 million Americans had Medicaid or CHIP coverage in any given month just before the pandemic hit in 2020, and millions more gained coverage during the pandemic when they lost their jobs and their health insurance. Total Medicaid spending in fiscal year 2022 topped $800 billion, making it one of the largest items in national and state budgets and the public safety net (Howard 2023; Kaiser Family Foundation 2023). Voter registration and turnout increased in Medicaid expansion states, which could boost political support for the program (Sanger-Katz 2018).

The long history of "welfare" narrowly defined—as it evolved through mothers' pensions, to Aid to Dependent Children, Aid to Families with Dependent Children,

Figure 4
Poverty Rates with and without Means-Tested Benefits (2017)



Source: Congressional Research Service (Falk, Carter, and Nicchitta 2021).
Note: Poverty rates are based on the Supplemental Poverty Measure. These benefits include TANF, SSI, EITC, SNAP, subsidized school meals, the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), public housing, Section 8 rental vouchers, and child care subsidies, but not Medicaid or CHIP.

and Temporary Assistance for Needy Families—does suggest that cash payment means-tested programs are politically vulnerable and will often prove inadequate. But welfare in this narrow sense never helped more than a small slice of the poverty population, and it has not been a major part of the public safety net over the last half century. Medicaid, the Earned Income Tax Credit, and Food Stamps (renamed in 2008 as SNAP, the Supplemental Nutrition Assistance Program) have become much more important, and their growth shows that programs for the poor can generate durable support, even as polarization among public officials intensified. The development of Food Stamps/SNAP, Supplemental Security Income, and the Earned Income Tax Credit shows that geographic inequalities in means-tested programs can be minimized. Politically, means-tested programs are more viable than they used to appear (Greenstein 2022).

These programs have reduced poverty, especially for children, families with children, and adults with disabilities, as shown in Figure 4. Without programs like the Earned Income Tax Credit and Supplemental Nutrition Assistance Program, child poverty rates in this country would double. The elderly, in contrast, rely more on Social Security and Medicare to stay out of poverty, so means-tested benefits have less of an impact (Falk, Carter, and Nicchitta 2021).[7] Figure 4 offers a snapshot of one year, but recent trends tell a similar story. Child poverty rates dropped by half between 1993 and 2019, a feat comparable to the decline of elder poverty in the 1960s and 1970s. Children in low-income households have benefitted enormously from expansions to the EITC. By 2019, the EITC was large enough to lift 2.4 million children out of poverty and lower the child poverty rate by more than three percentage points. The SNAP and SSI programs combined to lift 1.6 million children out of poverty, while the impact of TANF was much smaller (Thomson et al. 2022). Medicaid expansion has both reduced poverty and improved health outcomes (Remler, Korenman, and Hyson 2017; Currie and Chorniy 2021).

For those who define the public safety net as means-tested programs, recent decades offer some hope. Despite growing political polarization, the national government has been doing more to help the working poor and the disabled poor (Greenstein 2022). The government has also been spending more on food and medical care for people with low incomes, regardless of their employment status or ability to work, in part because these in-kind programs are supported by businesses that provide food and medical care. People with low incomes are not the only ones who benefit from SNAP and Medicaid (Olson 2010; Howard 2023).

For those who count social insurance programs as part of the safety net, Social Security is a major success story. Older Americans are much less likely to live in poverty than they were two generations ago, and Social Security is a big reason why. Unlike means-tested programs, the inclusiveness of Social Security and Medicare creates a large constituency of voters who care about these programs and can influence policy debates. There is, in short, no single formula for building a durable safety net—which may be good news considering how much remains to be done in the United States to reduce poverty, food insecurity, housing cost burdens, gaps in health insurance, and related social problems (Howard 2023).

References

Arnold, R. Douglas. 2022. *Fixing Social Security: The Politics of Reform in a Polarized Age.* Princeton University Press.

Béland, Daniel. 2005. *Social Security: History and Politics from the New Deal to the Privatization Debate.* University Press of Kansas.

Bell, Winifred. 1965. *Aid to Dependent Children.* Columbia University Press.

Berkowitz, Edward D. 1991. *America's Welfare State: From Roosevelt to Reagan.* Johns Hopkins University Press.

[7] As in Figure 3, the numbers in Figure 4 do not assume behavioral changes; they compare the incomes of various groups with and without means-tested benefits. Importantly, Figure 4 does not include the impact of Medicaid on poverty.

Berkowitz, Edward D., and Larry DeWitt. 2017. *The Other Welfare: Supplemental Security Income and U.S. Social Policy*. Cornell University Press.

Bertram, Eva. 2015. *The Workfare State: Public Assistance Politics from the New Deal to the New Democrats*. University of Pennsylvania Press.

Borelli, Christopher. 2019. "Reagan Used Her, the Country Hated Her. Decades Later, the Welfare Queen Refuses to Go Away." *Chicago Tribune*, June 10. https://www.chicagotribune.com/entertainment/ct-ent-welfare-queen-josh-levin-0610-story.html (accessed January 8, 2024).

Bosso, Christopher. 2023. *Why SNAP Works: A Political History—and Defense—of the Food Stamp Program*. University of California Press.

Brady, David. 2023. "Poverty, Not the Poor." *Science Advances* 9 (34): eadg1469.

Campbell, Andrea Louise. 2003. *How Policies Make Citizens: Senior Political Activism and the American Welfare State*. Princeton University Press.

Campbell, Andrea Louise, and Kimberly J. Morgan. 2005. "Financing the Welfare State: Elite Politics and the Decline of the Social Insurance Model in America." *Studies in American Political Development* 19 (2): 173–95.

Cates, Jerry R. 1983. *Insuring Inequality: Administrative Leadership in Social Security, 1935–54*. University of Michigan Press.

Center on Budget and Policy Priorities. 2022. "Policy Basics: Temporary Assistance for Needy Families." https://www.cbpp.org/sites/default/files/atoms/files/7-22-10tanf2.pdf (accessed January 9, 2024).

Center on Budget and Policy Priorities. 2024. "Policy Basics: Where Do Our Federal Tax Dollars Go?" Center on Budget and Policy Priorities, July 18. https://www.cbpp.org/research/federal-budget/where-do-our-federal-tax-dollars-go (accessed November 25, 2024).

Currie, Janet, and Anna Chorniy. 2021. "Medicaid and Child Health Insurance Program Improve Child Health and Reduce Poverty But Face Threats." *Academic Pediatrics* 21 (8): S146–53.

Dalaker, Joseph. 2022. *The Supplemental Poverty Measure: Its Core Concepts, Development, and Use*. Congressional Research Service.

Davies, Gareth. 1996. *From Opportunity to Entitlement: The Transformation and Decline of Great Society Liberalism*. University Press of Kansas.

Derthick, Martha. 1979. *Policymaking for Social Security*. Brookings Institution.

Falk, Gene. 2021. *Trends in TANF Receipt: Select Information for the Debate to Extend the Expanded Child Tax Credit*. Congressional Research Service.

Falk, Gene, Jameson A. Carter, and Isaac A. Nicchitta. 2021. *Need-Tested Benefits: Impact of Assistance on Poverty Experienced by Low-Income Families and Individuals*. Congressional Research Service.

Fishback, Price V. 2018. "Welfare Policy: The Public Safety Net in American Economic History." In *The Oxford Handbook of American Economic History*, Vol. 2, edited by Louis P. Cain, Price V. Fishback, and Paul W. Rhode, 447–74. Oxford University Press.

Fox, Cybelle. 2012. *Three Worlds of Relief: Race, Immigration, and the American Welfare State from the Progressive Era to the New Deal*. Princeton University Press.

Garland, David. 2016. *The Welfare State: A Very Short Introduction*. Oxford University Press.

Gilens, Martin. 1999. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. University of Chicago Press.

Goodwin, Joanne L. 1997. *Gender and the Politics of Welfare Reform: Mothers' Pensions in Chicago, 1911–1929*. University of Chicago Press.

Gordon, Linda. 1994. *Pitied But Not Entitled: Single Mothers and the History of Welfare, 1890–1935*. Harvard University Press.

Graetz, Michael J., and Jerry L. Mashaw. 1999. *True Security: Rethinking American Social Insurance*. Yale University Press.

Greenstein, Robert. 2022. "Targeting vs. Universalism, and Other Factors That Affect Social Programs' Political Strength and Durability." https://www.hamiltonproject.org/wp-content/uploads/2023/0½0220806_ES_THP_SocialPrograms_ExpandedEdition.pdf (accessed January 10, 2024).

Gwyn, Nick, and Jenna Gerry. 2023. "Unemployment Insurance System Unprepared for Another Recession." Center on Budget and Policy Priorities, April 18. https://www.cbpp.org/research/economy/unemployment-insurance-system-unprepared-for-another-recession (accessed December 5, 2024).

Haskins, Ron. 2006. *Work Over Welfare: The Inside Story of the 1996 Welfare Reform Law*. Brookings Institution Press.

Heclo, Hugh. 1986. "The Political Foundations of Antipoverty Policy." In *Fighting Poverty: What Works and*

What Doesn't, edited by Sheldon H. Danziger and Daniel H. Weinberg, 312–40. Harvard University Press.
Herndon, Ruth Wallis. 2001. *Unwelcome Americans: Living on the Margin in Early New England*. University of Pennsylvania Press.
Hiltzik, Michael. 2021. "Democrats Are Talking about Means-Testing More Social Programs. What a Terrible Idea." *Los Angeles Times,* October 21. https://www.latimes.com/business/story/2021-10-21/democrats-means-testing-social-programs (accessed January 5, 2024).
Howard, Christopher. 1992. "Sowing the Seeds of 'Welfare': The Transformation of Mothers' Pensions, 1900–1940." *Journal of Policy History* 4 (2): 188–226.
Howard, Christopher. 2002. "Workers' Compensation, Federalism, and the Heavy Hand of History." *Studies in American Political Development* 16 (1): 28–47.
Howard, Christopher. 2007. *The Welfare State Nobody Knows: Debunking Myths About U.S. Social Policy*. Princeton University Press.
Howard, Christopher. 2023. *Who Cares: The Social Safety Net in America*. Oxford University Press.
Jensen, Laura S. 2015. "Social Provision Before the Twentieth Century." In *The Oxford Handbook of U.S. Social Policy*, edited by Daniel Béland, Christopher Howard, and Kimberly J. Morgan, 23–40. Oxford University Press.
Joint Committee on Taxation. 2022. *Estimates of Federal Tax Expenditures for Fiscal Years 2022–2026*. JCX-22-22. Congress of United States.
Kaiser Family Foundation. 2023. *Total Medicaid Spending*. https://www.kff.org/medicaid/state-indicator/total-medicaid-spending/ (accessed January 9, 2024).
Katz, Michael B. 1986. *In the Shadow of the Poorhouse: A Social History of Welfare in America*. Basic Books.
Kaus, Mickey. 1986. "The Work Ethic State." *The New Republic* 195 (1): 22–33.
Kennedy, David M. 1999. *Freedom From Fear: The American People in Depression and War, 1929–1945*. Oxford University Press.
Kim, Christine, and Robert Rector. 2006. "Welfare Reform Turns Ten: Evidence Shows Reduced Dependence, Poverty." Heritage Foundation, August 1. https://www.heritage.org/welfare/report/welfare-reform-turns-ten-evidence-shows-reduced-dependence-poverty (accessed January 10, 2024).
Li, Zhe, and Paul S. Davies. 2022. *Income for the Population Aged 65 and Older: Evidence from the Health Retirement Study (HRS)*. Congressional Research Service.
Lieberman, Robert C. 1998. *Shifting the Color Line: Race and the American Welfare State*. Harvard University Press.
Lowell, Josephine Shaw. 1884. *Public Relief and Private Charity*. G. P. Putnam's Sons.
Lubove, Roy. 1968. *The Struggle for Social Security 1900–1935*. Harvard University Press.
McKernan, Signe-Mary, Caroline Ratcliffe, and Breno Braga. 2021. "The Effect of the US Safety Net on Material Hardship over Two Decades." *Journal of Public Economics* 197: 104403.
McGarry, Kathleen. 2013. "The Safety Net for the Elderly." In *Legacies of the War on Poverty*, edited by Martha J. Bailey and Sheldon Danziger, 179–205. Russell Sage Foundation.
Mead, Lawrence M. 1992. *The New Politics of Poverty: The Nonworking Poor in America*. Basic Books.
Mettler, Suzanne. 1998. *Dividing Citizens: Gender and Federalism in New Deal Public Policy*. Cornell University Press.
Moffitt, Robert A. 2013. "The Great Recession and the Social Safety Net." *Annals of the American Academy of Political and Social Science* 650 (1): 143–66.
Murray, Charles. 1984. *Losing Ground: American Social Policy, 1950–1980*. Basic Books.
Oberlander, Jonathan. 2003. *The Political Life of Medicare*. University of Chicago Press.
Olson, Laura Katz. 2010. *The Politics of Medicaid*. Columbia University Press.
Orloff, Ann Shola. 1991. "Gender in Early U.S. Social Policy." *Journal of Policy History* 3 (3): 249–81.
Patterson, James T. 1994. *America's Struggle Against Poverty, 1900–1994*. Harvard University Press.
Powell, Ernie. 2019. "Why America Is So Poor at Helping People Who Need Long-Term Health Care." *Salon,* December 13. https://www.salon.com/2019/12/13/why-america-is-so-poor-at-helping-people-who-need-long-term-health-care_partner/ (accessed January 12, 2024).
Quadagno, Jill. 1994. *The Color of Welfare: How Racism Undermined the War on Poverty*. Oxford University Press.
Rachidi, Angela, Matt Weidinger, and Scott Winship. 2022. "A Safety Net for the Future: Overcoming the Root Causes of Poverty." In *American Renewal: A Conservative Plan to Strengthen the Social Contract and Save the Country's Finances*, edited by Paul Ryan and Angela Rachidi, 132–58. American Enterprise Institute.

Radford, Gail. 1996. *Modern Housing for America: Policy Struggles in the New Deal Era*. University of Chicago Press.

Reagan, Ronald. 1986. "Radio Address to the Nation on Welfare Reform." February 15. https://www.reaganlibrary.gov/archives/speech/radio-address-nation-welfare-reform (accessed January 9, 2024).

Remler, Dahlia K., Sanders D. Korenman, and Rosemary T. Hyson. 2017. "Estimating the Effects of Health Insurance and Other Social Programs on Poverty Under the Affordable Care Act." *Health Affairs* 36 (10): 1828–37.

Romig, Kathleen. 2023. "Social Security Lifts More People Above the Poverty Line Than Any Other Program." Center on Budget and Policy Priorities, January 31. https://www.cbpp.org/research/social-security/social-security-keeps-22-million-americans-out-of-poverty-a-state-by-state (accessed January 14, 2024).

Romig, Kathleen, and Kathleen Bryant. 2021. "A National Paid Leave Program Would Help Workers, Families." Center on Budget and Policy Priorities, April 27. https://www.cbpp.org/research/economy/a-national-paid-leave-program-would-help-workers-families (accessed January 12, 2024).

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright.

Safire, William. 1981. "Safety Nets." *New York Times*, March 29. https://www.nytimes.com/1981/03/29/magazine/on-language-safety-nets.html (accessed January 5, 2024).

Sanger-Katz, Margot. 2018. "When Medicaid Expands, More People Vote." *New York Times*, November 8. https://www.nytimes.com/2018/11/08/upshot/medicaid-expansion-voting-increase.html (accessed January 11, 2024).

Scott, Dylan. 2019. "How to Build a Medicare-for-All Plan, Explained by Somebody Who's Thought about It for 20 Years." *Vox*, January 28. https://www.vox.com/policy-and-politics/2019/1/28/18192674/medicare-for-all-cost-jacob-hacker (accessed January 12, 2024).

Shrider, Emily A. 2024. *Poverty in the United States: 2023*. United States Census Bureau.

Singer, Audrey. 2004. "Welfare Reform and Immigrants: A Policy Review." In *Immigrants, Welfare Reform, and the Poverty of Policy*, edited by Philip Kretsedemans and Ana Aparicio, 21–34. Praeger.

Skocpol, Theda. 1991. "Targeting within Universalism: Politically Viable Policies to Combat Poverty in the United States." In *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson, 411–36. Brookings Institution Press.

Skocpol, Theda. 1992. *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States*. Harvard University Press.

Smith, David Barton. 2015. "Civil Rights and Medicare: Historical Convergence and Continuing Legacy." In *Medicare and Medicaid at 50: America's Entitlement Programs in the Age of Affordable Care*, edited by Alan B. Cohen, David C. Colby, Keith A. Wailoo, and Julian E. Zelizer, 21–38. Oxford University Press.

Social Security Administration. 2000. *Annual Statistical Supplement, 2000*. Social Security Administration, Office of Policy, Office of Research, Evaluation, and Statistics.

Social Security Administration. 2023. *SSI Annual Statistical Report, 2022*. SSSA Publication 13-11827.

Social Security Administration. n.d. "Text of the 1935 Social Security Act." https://www.ssa.gov/history/35actinx.html (accessed January 6, 2024).

Soule, Sarah A., and Yvonne Zylan. 1997. "Runaway Train? The Diffusion of State-Level Reform in ADC/AFDC Eligibility Requirements, 1950–1967." *American Journal of Sociology* 103 (3): 733–62.

Starr, Paul. 2013. *Remedy and Reaction: The Peculiar American Struggle over Health Care Reform*. Revised ed. Yale University Press.

Stone, Chad, and William Chen. 2014. "Introduction to Unemployment Insurance." Center on Budget and Policy Priorities, July 30. https://www.cbpp.org/research/introduction-to-unemployment-insurance (accessed January 12, 2024).

Tach, Laura, and Kathryn Edin. 2017. "The Social Safety Net After Welfare Reform: Recent Developments and Consequences for Household Dynamics." *Annual Review of Sociology* 43: 541–61.

Teles, Steven M. 1996. *Whose Welfare? AFDC and Elite Politics*. University Press of Kansas.

Thomson, Dana, Renee Ryberg, Kristen Harper, James Fuller, Katherine Paschall, Jody Franklin, and Lina Guzman. 2022. "Lessons From a Historic Decline in Child Poverty." ChildTrends, September 11. https://www.childtrends.org/publications/lessons-from-a-historic-decline-in-child-poverty (accessed January 12, 2024).

Trattner, Walter I. 1974. *From Poor Law to Welfare State: A History of Social Welfare in America*. Free Press.

US Bureau of the Census. 1964. *Statistical Abstract of the United States: 1964*. Government Printing Office.

US Bureau of the Census. 1972. *Statistical Abstract of the United States: 1972*. Government Printing Office.

US Bureau of the Census. 1982. *Statistical Abstract of the United States: 1982–83*. Government Printing Office.

US Bureau of the Census. 1994. *Statistical Abstract of the United States: 1994*. Government Printing Office.

US Bureau of the Census. 2000. *Statistical Abstract of the United States: 2000*. Government Printing Office.

US Census Bureau. 2023a. *Historical Poverty Tables: People and Families—1959 to 2022*. https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-people.html (accessed January 5, 2024).

US Census Bureau. 2023b. Poverty Thresholds. https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html (accessed on January 5, 2024).

US Census Office. 1881. *Statistics of the Population of the United States at the Tenth Census (June 1, 1880)*. Government Printing Office.

von Hoffman, Alexander. 2012. "History Lessons for Today's Housing Policy: The Politics of Low-Income Housing." *Housing Policy Debate* 22 (3): 321–76.

Ward, Deborah E. 2005. *The White Welfare State: The Racialization of U.S. Welfare Policy*. University of Michigan Press.

Weaver, R. Kent. 1988. *Automatic Government: The Politics of Indexation*. Brookings Institution Press.

Weaver, R. Kent. 2000. *Ending Welfare As We Know It*. Brookings Institution Press.

Whittaker, Julie M., and Katelin P. Isaacs. 2019. *Unemployment Insurance: Programs and Benefits*. Congressional Research Service.

Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

Witte, Edwin E. 1963. *The Development of the Social Security Act*. University of Wisconsin Press.

Woodbury, Stephen A. 2015. "Unemployment Insurance." In *The Oxford Handbook of U.S. Social Policy*, edited by Daniel Béland, Christopher Howard, and Kimberly J. Morgan, 471–90. Oxford University Press.

Zelizer, Julian E. 1998. *Taxing America: Wilbur D. Mills, Congress, and the State, 1945–1975*. Cambridge University Press.

Zollars, Cheryl, and Theda Skocpol. 1994. "Cultural Mythmaking as a Policy Tool: The Social Security Board and the Construction of a Social Citizenship of Self Interest." *Research on Democracy and Society* 2: 381–408.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 101–128

Did Welfare Reform End the Safety Net as We Knew It? The Record since 1996

Lucie Schmidt, Lara Shore-Sheppard, and Tara Watson

The set of supports for low-income populations in the United States, now commonly referred to as the "social safety net," is a complex system consisting of multiple programs, created at different times with different goals, administered and funded at different levels of government, with different benefit structures, and serving different (but partially overlapping) populations. Resources provided by the safety net have been credited with reducing material hardship and improving child outcomes, but programs have come under scrutiny because they may have unintended consequences for labor supply and family structure. Nearly three decades ago, the safety net experienced a watershed moment with the passage of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, commonly referred to as "welfare reform."

During the 1992 presidential campaign, President Bill Clinton often repeated a campaign promise to "end welfare as we know it." The vision included strengthening work requirements, limiting the duration of benefit receipt, and developing a broader set of supports (Soss and Schram 2006). After successful midterm elections for the Republicans in 1994, Clinton began to negotiate a compromise bill,

■ *Lucie Schmidt is Robert A. Woods Professor of Economics, Smith College, Northampton, Massachusetts. Lara Shore-Sheppard is Dean of the Faculty and Kimberly A. '96 and Robert R. '62 Henry Professor of Economics, Williams College, Williamstown, Massachusetts. Tara Watson is Director of the Center for Economic Security and Opportunity, John C. and Nancy D. Whitehead Chair, and Senior Fellow in Economic Studies, Brookings Institution, Washington, DC. All three authors are also Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Schmidt is the corresponding author at lschmidt@smith.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20231392.

eventually passing a welfare reform bill after two vetoes of earlier versions. The bill received about 75 percent of the vote in both the House and the Senate, with almost all of the "nay" votes from Democrats concerned that the legislation would reduce resources for low-income families. Several high-ranking officials in the Clinton administration resigned in protest, with one of them later calling the bill "the worst thing Bill Clinton has done" (Mitchell 1996; Edelman 1997).

The reform dismantled the traditional cash welfare system, in which low-income families received largely unconditional checks from the federal-state program called Aid to Families with Dependent Children. The program was renamed Temporary Assistance to Needy Families, and it included a lifetime limit on welfare payments of five years for any individual, requirements that states impose work or job-training requirements as a condition for receiving welfare payments, and flexibility for states to use block grant funding for cash transfers or in a variety other ways.

Both during the lead-up to the 1996 welfare reform act and since then, the animating axis of safety net policy and politics has remained the same; that is, a tension on one hand between reducing material hardship for low-income households, particularly for children in these families, and on the other hand promoting household labor market participation and self-sufficiency (Aizer, Hoynes, and Lleras-Muney 2022). Questions of who is expected to be self-sufficient and who is deserving of support remain politically contested. A related axis of debate in the US context pertains to federalism, most notably the degree of state flexibility and control in the design and implementation of safety net programs. The past quarter-century has seen considerable push-and-pull along these policy axes for major anti-poverty programs.

Though some feared the 1996 welfare reform would lead to the elimination of welfare, and that increased state flexibility would lead to a "race to the bottom" in terms of state benefit provision, the overall safety net is now much larger than in the past. But the safety net as it evolved is not simply a more expansive version of what it was before. In this essay, we consider how the safety net for low-income families and nonelderly adults without disabilities has evolved in the last three decades, along with the implications for who receives assistance, how much, and in what form.

In the first section of the paper, we describe the evolution of the major programs that are the core of our analysis: traditional cash transfers through the Aid to Families with Dependent Children (AFDC) program and its replacement Temporary Assistance for Needy Families (TANF), food assistance through the Supplemental Nutrition Assistance Program (SNAP, formerly known as Food Stamps), refundable tax credits through the Earned Income Tax Credit (EITC) and Child Tax Credit (CTC), and public health insurance through Medicaid and the Children's Health Insurance Program (CHIP).[1] Specifically, we characterize changes in the major

[1] This list of programs is by no means exhaustive. Other important means-tested programs that we do not discuss include Supplemental Security Income (SSI), which provides cash transfers for the elderly and people with disabilities; additional food programs such as the School Breakfast Program, the National School Lunch Program, and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC); and low-income housing programs. For more detail on these programs, see

means-tested US programs for the nonelderly, nondisabled population since 1996, and thus we do not discuss social insurance programs such as Unemployment Insurance and Social Security Disability Insurance, even though those programs improve resources available to families and can be considered part of the safety net (for a discussion of social insurance programs as part of the safety net, see the paper by Howard in this symposium).

In the next section, we consider the combined impacts of changes in these programs on the evolution of the safety net as a whole since 1996: coverage of more families and with more support; a shift from cash transfers towards tax credits and in-kind transfers; a shift toward married-parent families with children and the working poor, and away from other low-income households; a shift toward federal rather than state-level funding of the safety net; and other factors.[2] We then discuss differences in safety net access by race, ethnicity, and immigration status, and end by offering some conclusions.

Changes in Major US Safety Net Programs

To set the stage, Figure 1 shows expenditures for the six programs we analyze (in 2022 dollars) divided by total population from years 1965 through 2022 (for Medicaid, we show only expenditures on nonelderly recipients without disabilities, and for the Child Tax Credit, we show only the refundable portion). Spending on these programs grew substantially between 1965 and 1995, which helped to produce political momentum for the Personal Responsibility and Work Opportunity Reconciliation Act, the 1996 welfare reform bill. While we will sketch some of the earlier history of these programs, our main focus will be on changes since 1996. In Table 1, we offer a more detailed description of the key features of these programs, and online Appendix Table 1 summarizes major policy changes since 1996.

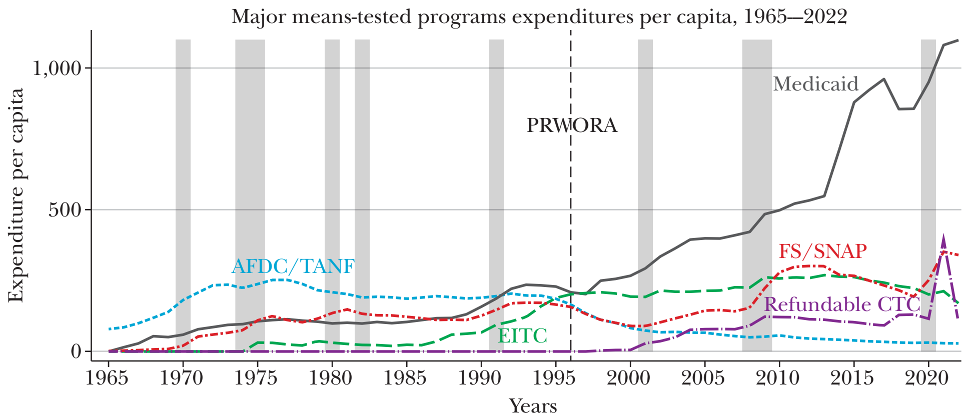
Aid to Families with Dependent Children/Temporary Assistance for Needy Families

The program colloquially referred to as "welfare," and previously called Aid to Dependent Children and then Aid to Families with Dependent Children, was created as part of the Social Security Act of 1935 in response to the Great Depression (Ziliak 2016). AFDC was funded through a federal match of state spending. It provided cash support targeted at single-parent families with very low incomes in the form of a monthly grant that was highest for families with no income.

Duggan, Kearney, and Rennane (2016) for Supplemental Security Income, Hoynes and Schanzenbach (2016) for food programs, and Collinson, Ellen, and Ludwig (2016) for housing programs.

[2] Other scholars have made similar and related points about the shifts in the US safety net, including Ben-Shalom, Moffitt, and Scholz (2012), Moffitt (2015, 2016), Hoynes and Schanzenbach (2018), and Han, Meyer, and Sullivan (2022). In this journal, Aizer, Hoynes, and Lleras-Muney (2022) document a shifting focus of economics research of safety net programs, away from papers focusing narrowly on potential work disincentives and toward papers that examine benefits to low-income households—including long-run benefits to children.

Figure 1
Expenditures per Capita, Major Cash, Near-Cash, and Health Insurance Programs, 1965–2022



Source: Expenditure data source varies by program and by year. See online Appendix.
Note: Figure shows program expenditures in inflation-adjusted 2022 dollars divided by population. AFDC/TANF dollars refer to basic monthly assistance and administrative spending. Medicaid dollars refers to spending on nondisabled and nonelderly populations where available, and imputed as a share of total Medicaid spending where unavailable. Refundable CTC refers to only the refundable portion of the Child Tax Credit, sometimes called the Additional Child Tax Credit. Gray bars indicate recessions, as dated by the National Bureau of Economic Research.

Throughout much of the program's history, the amount of the grant fell by nearly $1 for every $1 as earnings rose (Moffitt and Rangarajan 1991). Although the disincentives to work were very high, the program in the early period served mostly widows with children at a time when the social expectations of labor force participation by mothers were low. However, changing demographics meant that by the 1970s and into the 1980s, welfare program caseloads increasingly consisted of never-married women (about 21 percent in 1976 and 52 percent by 1992), who were often viewed as "less deserving" of support than the widows of the past (Katz 1989; US General Accounting Office 1994; Moffitt 2015).

As Figure 1 shows, per capita spending on the Aid to Families with Dependent Children program peaked around 1977 and remained at this higher level into the 1990s. In addition, both implicit and explicit forms of racial discrimination that had existed in the program since its inception began to decline as a result of the civil and welfare rights movements, the introduction of community legal services, and a number of Supreme Court decisions that struck down state-level eligibility restrictions on welfare benefits, like restrictions on cohabitation or certain residency requirements (for discussion, see Floyd et al. 2021; Cunningham and Goodman-Bacon 2025; Fuller 2023). As a result, welfare caseloads of women of color increased, further fueling a racially coded discussion of "deservingness." This, coupled with changing American norms around mothers in the labor force, weakened public support for the program.

Table 1
Programs and Details

Program name	*Type of program and level of funding/administration*	*Eligible population*	*2019 income eligibility threshold*	*2019 annual maximum benefit level*
Temporary Assistance for Needy Families (TANF)	Cash; funded through federal block grant, states administer with some federal guidelines.	Families with children are eligible; most generous for single-parent families.	Ranges from ~$3K to ~$27K for a single-parent family of three, depending on state.	Ranges from ~$2K to ~$13K for a single-parent family of three, depending on state.
Earned Income Tax Credit (EITC)	Refundable tax credit; federal, 28 states have their own additional credit as of 2019.	Tax filers with earned income are eligible; most generous for those with children.	$47K for a single-parent family of three.	Max federal benefit is ~$6K for single-parent family, plus additional state benefit depending on state.
Child Tax Credit (CTC)	Partially refundable tax credit; federal	Tax filers with children are eligible if they have at least $2,500 in earned income.	Exceeds $200K for single-parent family.	Refundable portion of tax credit is up to $1,400 per child.
Supplemental Nutrition Assistance Program (SNAP)	In-kind food; federal	Households	~$27K for a family of three.	~$6K for a family of three.
Medicaid and Children's Health Insurance Program (CHIP)	In-kind health insurance; funded jointly at federal and state level.	Children, parents, elderly and disabled are categorically eligible. In some states, anyone who meets state-set income cutoffs is eligible; in other states, nonelderly adults without disabilities or children remain ineligible regardless of income.	$9K to $29K for parents; $37K to $86K for children, depending on state.	

Source: Authors' creation. For additional detail on these programs, see Ziliak (2016) on AFDC/TANF, Nichols and Rothstein (2016) on tax credits, Hoynes and Schanzenbach (2016) on SNAP, and Buchmueller, Ham, and Shore-Sheppard (2016) on Medicaid and CHIP.

The 1996 Personal Responsibility and Work Opportunity Reconciliation Act replaced AFDC with Temporary Assistance to Needy Families (TANF). The revised law included stringent eligibility criteria including requirements that recipients work, time limits on lifetime benefit receipt, and sanctions to remove recipients for noncompliance with various program rules (for a contemporaneous discussion of the welfare reform legislation in this journal, see Blank 1997). Low-income single-parent families no longer had an entitlement to cash support.

The welfare reform act also restructured the program into a state block grant. States gained a great deal of discretion to set eligibility requirements and benefit amounts, and to spend funds on poverty-related programming other than cash transfers such as child care, state-level earned income tax credits, and one-time diversion payments; indeed, Temporary Assistance for Needy Families dollars can be spent on programming aimed at promoting marriage and reducing single parent families, even if those dollars are not focused on low-income individuals. The share of TANF dollars being spent on cash assistance fell, from between two-thirds to three-quarters under Aid to Families with Dependent Children to 36 percent by 2012 (Ziliak 2016). Benefit eligibility rules and the cultural orientation of welfare offices changed to discourage welfare participation and encourage work. States continue to target support towards very low-income families with children; the median state income limit for program eligibility is 48 percent of the federal poverty level (Falk 2014). Over the subsequent decades, the real value of the state block grants, which were fixed in nominal terms, eroded.

Both before and after welfare reform, Aid to Families with Dependent Children or Temporary Assistance for Needy Families benefit amounts were insufficient to lift nonworking families above the poverty line, even among those receiving assistance. For example, in 1994 the maximum monthly benefit for a family of three ranged from \$120 (about \$255 in 2024 dollars) in Mississippi to \$923 (\$1,959 in 2024 dollars) in Alaska, with the median state of Maryland offering \$366 (\$777) (National Research Council 1995). The 1994 continental US poverty line for the same family was \$12,320 annually, or \$1,027 monthly (\$2,179 in 2024 dollars). After reform, states continued to experiment with and in some cases expand the generosity of the TANF program. For example, 21 states and the District of Columbia raised TANF benefit levels between July 2022 and July 2023 (Azevedo-McCaffrey and Aguas 2024). Nevertheless, the real value of maximum benefits eroded in most states in the quarter-century after reform. In 2024, monthly maximum benefits for a family of three ranged from \$204 in Arkansas to \$1,370 in Minnesota, with a median of \$552 in Nebraska (Chatfield 2024). With the 2024 monthly poverty line for a family of three at \$2,152, TANF remains insufficient as a sole source of income to bring recipient families out of poverty.

Total expenditures on cash welfare fell sharply after welfare reform (as shown in Figure 1), mainly reflecting a sharp decline in participation. In 1996, 68 percent of families with children living under the federal poverty line received cash transfers through Temporary Assistance for Needy Families; by 2019, only 23 percent of these families received this form of cash assistance (Pavetti, Safawi, and Trisi 2021). Cash assistance caseloads fell from more than five million families in 1994 to around one million in 2023 (Congressional Research Service 2024). Today, most families with children below the poverty line do not receive TANF payments, and TANF is a shrinking component of the safety net.

Earned Income Tax Credit

The federal Earned Income Tax Credit was introduced in 1975, in part as a response to emerging concerns that Aid to Families with Dependent Children and other programs had created work disincentives (Nichols and Rothstein 2016; Bastian

2020). Unlike the structure of AFDC/TANF, where the highest benefits are directed at the lowest earners, the Earned Income Tax Credit is only available to those with earnings. The incentives with respect to marriage and work depend on the potential earnings of family members, but in general the EITC structure encourages entry into the labor market for at least one parent, particularly relative to the structure of the old AFDC system. In 2024, for example, a single parent with two children with earnings up to $17,400 was eligible for a credit of 40 percent of the amount earned, topping out at $6,960 (Tax Policy Center 2024a). (For comparison, this is very similar to the maximum amount that could be received for a TANF-participating family of three in the median state, though the fraction of families able to access TANF benefits is much lower.) The EITC is stable as income rises from $17,400 to $22,720, and then phases out at a rate of 21.06 percent; these figures are all adjusted annually to account for inflation. Although a small benefit is available to childless tax filing units, the bulk of EITC support goes to single- and married-parent families with children.

Importantly, the Earned Income Tax Credit is "refundable," meaning that those with tax liability smaller than the credit amount may receive the excess credit as a refund from the government. The result is a lump-sum transfer that typically arrives in the first quarter of the year, unlike the monthly transfer found in traditional welfare programs.

The Earned Income Tax Credit has been expanded several times with bipartisan support. Figure 1 shows substantial increases in federal expenditures on the EITC in the mid- to late-1990s due to a major expansion in 1993, and another increase with the expansion of the credit for families with three or more children as part of the American Recovery and Reinvestment Act in 2009. During the 2021 tax year in the pandemic, there was a temporary expansion of the federal EITC for workers without dependents, who typically qualify only at very low earnings levels and only for small benefits. In addition, 31 states plus Washington, DC, have opted to supplement the federal EITC with a state-level EITC, most often structured as a percentage of the federal EITC (Tax Policy Center 2024b).

Child Tax Credit

The Child Tax Credit was introduced in 1997 as a way of using the tax code to assist families with children. Initially this credit was nonrefundable, limiting its use by the lowest-income families, but it was made partially refundable in 2001, and then significantly expanded as part of the Tax Cuts and Jobs Act of 2017. In 2024, the CTC provides up to $2000 per child, of which $1700 is refundable. The CTC is not available to those without earnings, phasing in beginning at earnings of $2500. It does not phase out until income levels exceed $200,000 per year (in 2024), making it available to most families with children. The expenditures shown in Figure 1 only reflect the refundable portion of the program (sometimes known as the Additional Child Tax Credit or ACTC), which is most relevant for low-income families. During the pandemic, the federal government substantially expanded the per-child credit to $3000 or $3600 depending on age, made the full amount available to families without earnings, and provided a portion of the benefit monthly rather than as a lump sum. In response to the expiration of this temporary Child Tax Credit, a handful of states have recently introduced their own versions of the credit.

These federal and state tax changes have boosted the cash resources available to the working poor and near-poor. Whereas in 1994 a single-parent family of three without earnings could have received up to $9,321 annually (in 2024 dollars) in Aid to Families with Dependent Children benefits in the median state, much less would have been available if the parent worked. Today, Temporary Assistance for Needy Families benefits are typically lower and available to far fewer families, but a working single-parent family of three in 2024 can receive up to $10,360 in refundable federal EITC/CTC in addition to their earnings. Importantly, federal EITC and CTC benefits are only available to families with earnings, highlighting the increased work incentives embedded in the current system and the reallocation of safety net resources towards the working poor.

Supplemental Nutrition Assistance Program

Changes in cash support have been accompanied by expansions in in-kind transfers. One of the largest among these is the Supplemental Nutrition Assistance Program. Beginning as the Food Stamp program and renamed SNAP in 2008, the program was launched as a pilot program in 1961 and was operating in all counties by 1975 (Hoynes and Schanzenbach 2009, 2016). It provides monthly benefits on an "Electronic Benefits Transfer" card which can be used only to purchase food for home preparation. SNAP operates as a near-cash program because benefit amounts are lower than typical grocery expenditures for many families, thus freeing up income that can be spent for other purposes, although the restriction to food expenditures may assuage concerns about misuse of resources (Campbell and Gaddis 2017).[3]

The Supplemental Nutrition Assistance Program is federally-funded and federally-directed. Households with gross income below 130 percent of the federal poverty line (below $32,328 for a family of three in 2024) and income net of "allowable deductions" below 100 percent of the federal poverty line are eligible (US Department of Agriculture 2024). (Allowable deductions include a standard deduction, a percentage of earned income, and certain expenses.) Recipients without income receive the maximum benefit amount ($766 for a family of three in 2024), with the benefit amount eventually falling by roughly 30 cents on the dollar as income rises. The maximum benefit is determined annually based on the US Department of Agriculture's "Thrifty Food Plan," the estimated cost of groceries needed to provide a healthy but low-cost diet, and provides the same level of benefits across the continental United States, with benefits higher in Alaska and Hawaii (US Department of Agriculture 2025).

[3] There is some debate in the literature about whether recipients spend SNAP in the same way as cash (for example, Hoynes and Schanzenbach 2009; Beatty and Tuttle 2015; and Hastings and Schapiro 2018).

Unlike Temporary Assistance to Needy Families, which is limited to families with children, the Supplemental Nutrition Assistance Program is available more broadly to most income-eligible individuals, although some groups face additional restrictions on benefit receipt. In particular, as a result of the welfare reform act of 1996, working-age able-bodied adults without dependents ("ABAWDs") face work requirements (Bauer and East 2023). States have some discretion over the application of work requirements, as well as enrollment practices and other features of the program. In addition, states indirectly influence SNAP benefits because state-directed TANF transfers are considered in the income eligibility determination—that is, those in states with less generous TANF programs receive higher SNAP amounts.

As seen in Figure 1, Supplemental Nutrition Assistance Program expenditures were comparable to those of Aid to Families with Dependent Children around the time of welfare reform, and fell with Temporary Assistance for Needy Families expenditures in the years immediately after 1996. However, while TANF expenditures continued to fall, SNAP expenditures have risen in recent decades. In the early 2000s, states responded to the decline in SNAP participation by experimenting with ways to facilitate enrollment. There were also temporary expansions to benefit generosity during the Great Recession. During the COVID era, emergency allotments increased benefit amounts by $95 or more per month, bringing those qualifying for low benefit amounts up to the maximum level. The Thrifty Food Plan was revamped in 2023, keeping benefits elevated even as COVID-related expansions phased out. Today, with higher benefit amounts and many more participants (there are 22 million SNAP households in 2024 compared to about one million TANF families), SNAP spending exceeds $100 billion annually and dwarfs TANF cash assistance spending by a factor of ten (Shivastava and Manansala 2024; US Department of Agriculture 2024).

Medicaid and the Children's Health Insurance Program

Medicaid provides public health insurance for low-income individuals. Upon its creation in 1965, Medicaid coverage was limited to actual or potential recipients of certain cash assistance programs, so that among low-income families, only those receiving Aid to Families with Dependent Children also received Medicaid.[4] Beginning in the 1980s, the program gradually expanded to address concerns over lack of health insurance among certain groups of low-income people. These expansions initially focused on pregnant women, infants, and young children, extending Medicaid eligibility to members of these groups with higher incomes than the AFDC income limits.

The welfare reform bill in 1996 decoupled Medicaid from the new Temporary Assistance for Needy Families program. Instead, Medicaid eligibility became determined by Medicaid-specific income criteria, although the categorical requirement of being a child, a parent, or pregnant remained. In 1997, the Balanced Budget

[4] In addition to members of families with a single parent, the other categorically eligible groups for Medicaid were low-income individuals with disabilities and the elderly.

Act established the State Children's Health Insurance Program, originally known as SCHIP but renamed CHIP in 2009, which allowed states to increase the income eligibility limits for health insurance for children to even higher levels.

The Patient Protection and Affordable Care Act of 2010 sought to expand Medicaid to cover all individuals with incomes below 138 percent of the federal poverty line. However, the US Supreme Court ruled in *National Federation of Independent Business v. Sebelius* (567 US 519 [2012]) that, given the federal-state structure of the law, states could not be required to expand Medicaid. As a result, almost half of the states did not expand Medicaid by the originally intended 2014 implementation date, but as of 2024, 41 states had done so.

During the COVID-19 pandemic, the federal government increased the share of Medicaid dollars coming from the federal government to the states, conditional on states maintaining continuous coverage for beneficiaries. This policy temporarily eliminated eligibility redeterminations and disenrollments, allowing many recipients to access the program for longer than they otherwise would have (Dague and Ukert 2023). The enhanced federal match and continuous coverage requirement came to an end in 2023.

Figure 1 shows the striking rise in per-capita Medicaid expenditures (excluding Medicaid spending on the elderly and those with disabilities); specifically, Medicaid spending per person was similar to the Earned Income Tax Credit and to Aid to Families with Dependent Children circa 1995, but while the AFDC/TANF level has declined and the EITC has remained fairly stable since then, Medicaid spending per capita has roughly tripled. The higher spending reflects growth in health care costs that consistently exceeds price growth of other goods and services as well as expansions in eligibility that led to growth in the number of participants in the program. There was a substantial rise in expenditures accompanying the 2009 Affordable Care Act legislation, and another increase during the COVID-19 pandemic.

Spillovers

Although we have described the key safety net programs individually, there are important spillovers across programs. For example, income from Temporary Assistance to Needy Families is counted towards the determination of eligibility for the Supplemental Nutrition Assistance Program, so changes to TANF generosity directly affect SNAP benefit levels. In addition, to the extent that rules for a given program encourage or discourage labor supply, eligibility for other programs may be impacted indirectly.

There can also be spillover effects stemming from administrative burdens, barriers to participation stemming from bureaucratic processes rather than explicit eligibility rules (as discussed by Herd and Moynihan 2018, and also in this symposium). Enrolling in one safety net program may lead to knowledge about or facilitate enrollment in others. For example, past Medicaid expansions have spurred enrollment in SNAP, presumably by reducing the marginal cost of applying (for example, Baicker et al. 2014; Burney, Boehm, and Lopez 2021; Schmidt, Shore-Sheppard and Watson 2021a; Cha and Escarce 2022).

Substitution across programs also occurs, often in ways that allow states to substitute away from state-funded benefits towards federally-funded benefits. For example, the introduction of a federal program providing cash assistance for individuals with disabilities, Supplemental Security Income (SSI), led to reductions in Aid to Families with Dependent Children caseloads that were funded by a federal-state match (Goodman-Bacon and Schmidt 2020); similarly, reductions in access to AFDC or other state cash programs led to increases in SSI (Bound, Kossoudji, and Ricart-Moes 1998; Kubik 2003, Schmidt and Sevak 2004).

In sum, the safety net is best considered holistically as a system. The populations served are overlapping (Schmidt, Shore-Sheppard, and Watson 2021b), participation in one program can facilitate or inhibit participation in another, and gaps left by one program may lead to policy responses in another. In the next section, we consider some main themes of the evolution of the safety net as a whole since the 1996 welfare reform.

Key Implications of Program Changes

We identify six key themes that characterize changes in the overall structure of the safety net for nonelderly, nondisabled individuals since the passage of the welfare reform act of 1996.

1) The safety net has substantially expanded in generosity since 1996. Around the time of the welfare reform act of 1996, critics feared that the safety net would all but disappear in the subsequent years. While concerns about the erosion of traditional cash transfers were well-founded, expansions in tax credit programs and the Supplemental Nutrition Assistance Program more than offset these post-1996 losses.

In 1995, combined real expenditures on the major federal cash and near-cash programs we consider hit their highest level up to that date, $543 per capita in 2022 dollars. Partly due to the 1996 welfare reform and partly due to a strong economy in the late 1990s, real per capita expenditures on these programs had fallen by 31 percent by the year 2000. It took until 2009 for real per capita expenditures to surpass their 1995 levels. Although real expenditures today are lower than during the Great Recession and the pandemic era, the 2022 per capita levels at $643 are 19 percent higher than in 1995. Moreover, public health insurance coverage became much more widespread post-1996, with coverage for some groups, particularly children, extending higher up the income distribution, and other groups becoming eligible. Other smaller programs targeted at low-income populations also expanded their reach during this time period, including energy assistance, school breakfast and lunch programs, and others (Perl 2018; Congressional Research Service 2021, 2023).

These expenditure patterns (apparent in Figure 1) stem from evolving program characteristics, as well as economic and demographic shifts. To isolate the changes to program rules, we use a multi-program safety net calculator developed in earlier work (Schmidt, Shore-Sheppard, and Watson 2016, 2023) to impute eligibility and potential benefit levels for Temporary Assistance to Needy Families, the Supplemental Nutrition Assistance Program, the federal and state Earned Income Tax Credit and the refundable portion of the federal Child Tax Credit, accounting for policy-induced interactions between programs. We generate measures of program generosity assuming families take full advantage of their eligibility. However, it is important to note that actual take-up rates (participation rates among eligible populations) are substantially less than 100 percent, and vary by program and by individual characteristics (for reviews, see Currie 2006; Ko and Moffitt 2022). Also, we ignore TANF work requirements in our measure of generosity because they are inconsistently applied (Hahn, Kassabian, and Zedlewski 2012; Falk 2023).

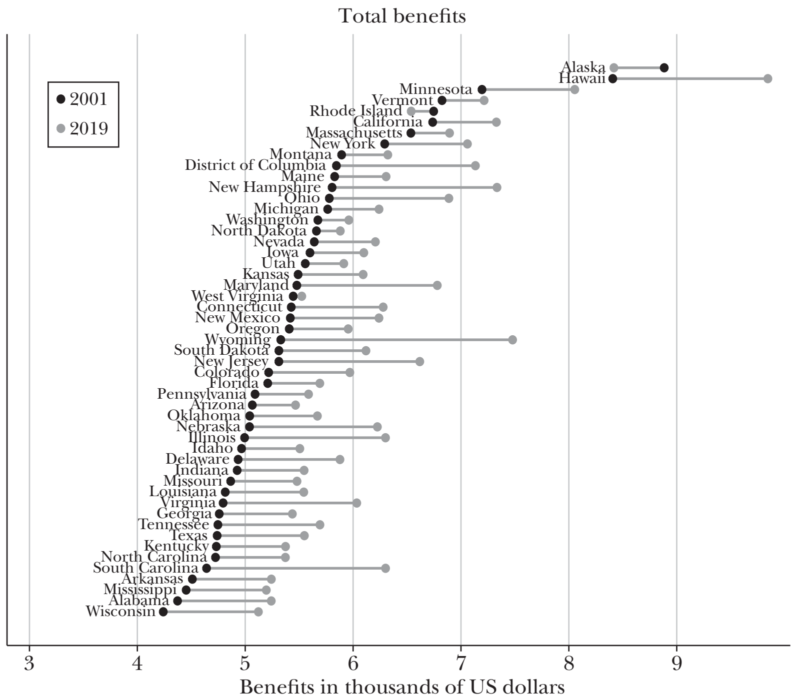
We focus on a subgroup that has traditionally been a key focus of the social safety net: single-parent families. Single-parent families have lower incomes on average than two-parent families, and some programs offer more generous benefits for single parents. Using the calculator, we select a simulated sample (using an approach pioneered by Currie and Gruber 1996) of nondisabled, nonimmigrant single-parent families that is representative of the national population of such families over the 2010–2019 period. We apply program rules for each simulated state and year to impute benefits for this consistent simulated set of families based on observed inflation-adjusted earnings and demographic information.

The benefits available on average in each state and year are then used to create an index, which represents inflation-adjusted annual benefits available to a typical single-parent family, assuming they fully accessed available major cash and near-cash benefits (Temporary Assistance for Needy Families, refundable tax credits, and the Supplemental Nutrition Assistance Program). Differences in the generosity index across states and over time stem only from differences in program rules, not demographic characteristics or economic conditions.

Figure 2 documents the generosity index for each state measured in 2001 and 2019 (using 2019 as an end point to avoid capturing the temporary safety net changes related to the pandemic). Nearly every state showed an expansion of cash and near-cash benefits available to single-parent families between 2001 and 2019, although the changes were not uniform across states. Aggregating to the national level using state population weights, we see that the generosity index rose by 14 percent, from $5,370 in 2001 to $6,112 in 2019.

2) Since 1996, support for low-income households has shifted away from direct monthly cash transfers towards tax credits and in-kind transfers. The composition of safety net expenditures has changed substantially since 1996. Traditional cash assistance has declined. Tax credits and food assistance have grown substantially, and in 2019 represented 95 percent of the total for the combined four cash- and near-cash key programs, compared to 66 percent in 1995. Even more striking is the increase in expenditures on Medicaid and the Children's Health Insurance Program, which now eclipse expenditures on all other major programs. With the increased flexibility given to states in the 1996 welfare reform, even the Temporary Assistance for Needy Families program has become reoriented away from cash support towards other types of spending.

Figure 2
Simulated Safety Net Benefits by State, 2001 and 2019



Source: Authors' calculations using rules for each program and their interactions.
Note: Simulated using a pooled consistent sample of single-parent, nondisabled, nonimmigrant families in the 2010–2019 CPS ASEC. Data points represent the average imputed benefits for this set of families as if they lived in a given state and year and they fully accessed cash- and near-cash benefits (TANF, federal EITC, state EITC refundable federal CTC, or SNAP) for which they are eligible given observed earnings and family structure. Reported in inflation-adjusted 2022 dollars. Simulation does not model work requirements or incomplete take-up.

A move towards in-kind rather than cash benefits brings inherent tradeoffs. Standard economic reasoning would suggest that families are better poised to allocate cash resources, and low-income families have needs that cannot be met with in-kind transfers. But some arguments do support an in-kind approach. For example, safety net benefits are sometimes provided to address multiple policy goals, such as Medicaid addressing an uninsurance problem while also boosting resources at the bottom of the income distribution. In addition, the use of in-kind benefits may quell fears of misuse of resources, thereby garnering more public support for redistribution.

The use of tax credits as opposed to cash transfers also presents trade-offs. By design, the federal Earned Income Tax Credit and Child Tax Credit do not help those families with zero earnings, which also limits the degree to which the programs can act as an automatic stabilizer (Bitler, Hoynes, and Schanzenbach 2020). However, the administrative burden of filing an annual tax return is likely lower than the onerous process of applying for and maintaining welfare benefits, and interviews suggest that recipients of the EITC greatly value the sense of dignity and reduced stigma of the tax credit relative to traditional welfare benefits (Halpern-Meekin et al. 2015). Tax benefits typically arrive annually rather than monthly, which may affect how they are spent and the degree to which they alleviate material hardship. In fact, qualitative evidence suggests that recipients strongly prefer the lump-sum option (Romich and Weisner 2000), and that four of every ten dollars of refund are invested or saved (Halpern-Meekin et al. 2015).

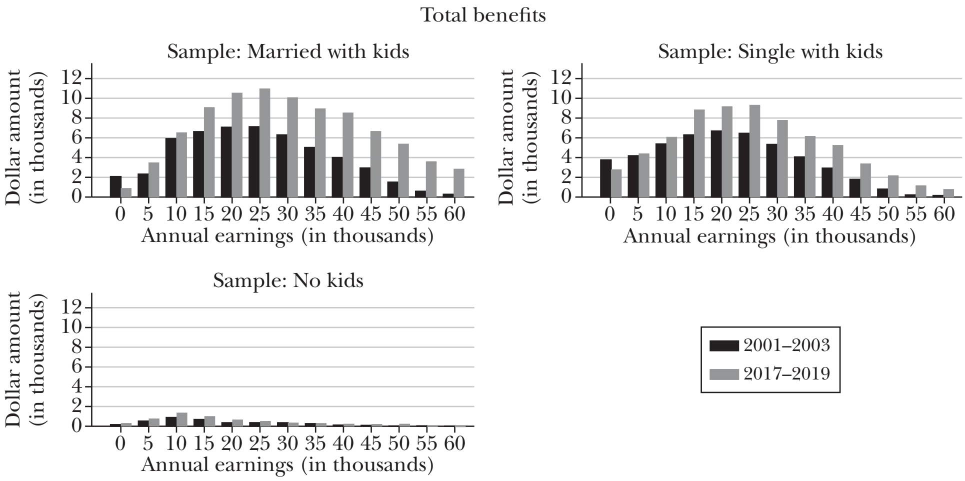
3) Safety net benefits, especially cash supports, have shifted away from families without earnings towards the working poor and near-poor. The 1996 welfare reform was fueled in large part by concerns that the system, which had very high implicit marginal tax rates at that time, disincentivized work. Thus, the Temporary Assistance to Needy Families program includes explicit work requirements and, in some states, revamping of benefit formulas to reduce implicit marginal tax rates (Matsudaira and Blank 2014). The gradual erosion of real welfare benefits over time and changes in welfare office culture also discouraged participation in the program. In addition, welfare reform's tightening of work requirements means that adults without dependents are effectively ineligible for the Supplemental Nutrition Assistance Program unless they work. The expanded tax credit programs, the Earned Income Tax Credit and the Child Tax Credit, both are contingent on a family having earned income, and credit amounts increase over the lower part of the income range (Hoynes and Patel 2018). As a result, excepting families who may be eligible for disability support programs, cash and near-cash support from the government is now highest at a modest earnings level rather than at zero earnings.

As part and parcel of shifting benefits away from zero-earnings families, the new structure of benefits creates stronger incentives for entry into the labor market than in the past. On the other hand, disincentives to work additional hours, or for secondary earners to participate in market work may have been exacerbated by recent reforms (Kearney and Turner 2013).

Figure 3a uses data from the Current Population Survey Annual Social and Economic Supplement (ASEC) to illustrate the total cash and near-cash safety net benefits reported (in 2022 dollars) by families of different types at various places throughout the earnings distribution for two periods: pooled 2001–2003 and pooled 2017–2019 (US Bureau of Census 2001–2003, 2017–2019). We focus on family units that do not contain disabled, elderly, or foreign-born individuals. Separate panels show benefits for married parents, single parents, and adults with no dependent children. Total benefits include self-reported benefits from Temporary Assistance for Needy Families and the Supplemental Nutrition Assistance Program, and imputed benefits from the federal and state Earned Income Tax Credit and the refundable portion of the Child Tax Credit.[5] Consistent with the notion that

[5] Program participation and benefits have been shown to be underreported in survey data, and this underreporting is worsening over time (for discussion in this journal, Meyer, Mok, and Sullivan 2015; see

Figure 3a
Total Safety Net Benefits Reported by Family Type and Place in Earnings Distribution, 2001 and 2019



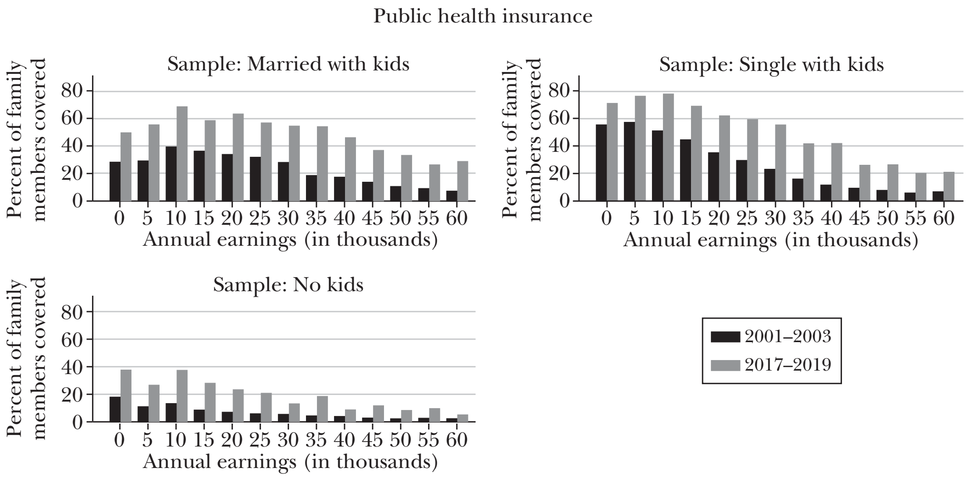
Source: CPS ASEC and authors' calculations.
Note: 2022 dollars. Sample is nondisabled, nonimmigrant families with at least one working age individual and no elderly family members in the 2001–2003 or 2017–2019 CPS ASEC. Annual earnings intervals show 0 for those earning $0, 5K for $1–$4,999, 10K for $5,000–$9,999, and so on. Total benefits include reported SNAP and TANF and imputed federal EITC, state EITC, and refundable federal CTC using TAXSIM.

safety net generosity has increased overall, families with children are receiving more benefits in 2017–2019 than they received in 2001–2003 at almost all points in the earnings distribution. However, the real value of reported cash and near-cash support has declined in the zero earnings category. The highest benefit levels and the most substantial increases over time are reported by families with incomes in the $15,000 to $45,000 range. This change reflects the growing importance of benefits that are conditioned on work, and a resulting reallocation of resources towards families with earned income.

By contrast, Figure 3b, which uses the same samples and graphs the average percentage of family members reporting public health insurance coverage, shows that public health insurance increased for all groups of families with earnings below $60,000. While reported public health insurance coverage rates increased the most in the same range of incomes that also saw growth in cash and near-cash benefits,

also Fox et al. 2017; Meyer and Mittag 2019; Meyer, Mittag, and Goerge 2022). SNAP and TANF benefits are therefore likely higher than suggested by the Annual Social and Economic Supplement survey. Because the EITC and CTC are not reported to the ASEC, we impute the Earned Income Tax Credit and the refundable portion of the Child Tax Credit based on earnings using the National Bureau of Economic Research TAXSIM model (Feenberg and Coutts 1993, 2024), and therefore implicitly assume full take-up of tax credits. Jones and Ziliak (2022) link the ASEC to IRS tax data, and find that the tax simulators overestimate the number of people lifted from poverty by the EITC.

Figure 3b
Public Health Insurance Reported by Family Type and Place in Earnings Distribution, 2001 and 2019



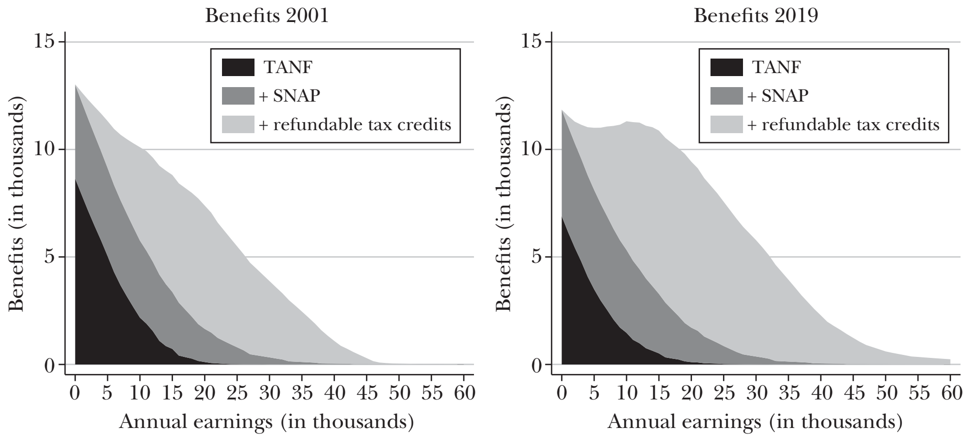
Source: CPS ASEC and authors' calculations.
Note: See Figure 3a note. Public health insurance includes one of the following: Medicaid, Medicare, CHIP, Indian health service, and other government health care.

health insurance also became more prevalent for non-earners. The expansion of public health insurance over this time period represents a major resource transfer towards low-income families.

To further assess differences in benefit amounts across the income distribution, we return to the simulated sample of single-parent families described above. Here, we assign to the constant simulated sample of families different earnings amounts in $1,000 increments; determine benefits according to program rules by state, year, and earnings; and aggregate to the national level, weighting by state population. Figure 4 illustrates the simulated benefits (in 2022 dollars) for Temporary Assistance for Needy Families, the Supplemental Nutrition Assistance Program, and the refundable tax credits available on paper to single-parent families by earnings level, for 2001 and 2019. Comparing the 2019 distribution of benefits to that seen in 2001, TANF has become a less substantial part of the total benefit package for families, particularly at zero earnings. This conclusion would be further strengthened if one accounted for the increasing difficulty non-earners have accessing TANF due to work requirements, which is not reflected in the generosity index presented in the figure.

The simulation in Figure 4 also shows that the Supplemental Nutrition Assistance Program expansions help to compensate for Temporary Assistance for Needy Families declines. This occurs both because reduced TANF lowers income and mechanically leads to higher SNAP support levels, but also because SNAP generosity expanded over this time period. In addition, Figure 4 illustrates the significant

Figure 4
Simulated Benefits for TANF, SNAP, and Refundable Tax Credits by Place in Earnings Distribution, 2001 and 2019



Source: Authors' creation.
Note: 2022 dollars. Simulated based on a consistent sample of single-parent, nondisabled, nonimmigrant families using observed family structure from the 2019 CPS ASEC. Data points represent the national average of imputed benefits for this set of families as if they lived in a given state and year, if they had a simulated earnings amount, and if they fully accessed cash- and near-cash benefits (TANF, federal EITC, state EITC refundable federal CTC, or SNAP) for which they are eligible given simulated earnings and observed family structure. Reported in inflation-adjusted 2022 dollars. Simulation does not model work requirements or incomplete take-up.

expansions of the refundable tax credits by 2019, reflecting both increased benefit amounts to families that were eligible in 2001 as well as expansions to families that are somewhat higher in the earnings distribution.

The results shown in Figure 4, which simulate potential benefits for a consistent set of families assuming full take-up of benefits at each earnings level, are not directly comparable to those shown in Figure 3a, which instead considers reported benefits for families at actual earnings levels. Thus, the much lower level of benefits suggested by Figure 3a reflects a combination of incomplete take-up, under-reporting, and compositional differences between the empirical sample and the simulated sample. Nevertheless, both approaches reach a similar conclusion: benefits available to single parents without earnings—presumably those with the greatest material need—have declined over recent decades. The simulated approach suggests that the cash and food benefit package potentially available to a typical single-parent family with no earnings fell from $13,026 to $11,848 in inflation-adjusted dollars. Separate analysis (not shown) suggests that this decline in resources to families without labor income is evident in almost every state.

When we add the benefits from the cash and near-cash programs together as in Figure 4, it is clear that incentives for labor market entry have increased substantially between 2001 and 2019, even separate from explicit Temporary Assistance for Needy Families work requirements, and that the implied marginal tax rate at very low levels of earnings has declined substantially. Around the time of welfare reform, a single mother entering the labor market and earning $15,000 would expect to lose more than $4,000 in cash and food support, with tax credits only partially offsetting the near-total loss of TANF. By 2019, she would lose fewer than $1,000 in benefits in the same scenario. The pattern of increased incentives for work over time is common across states (results not shown).

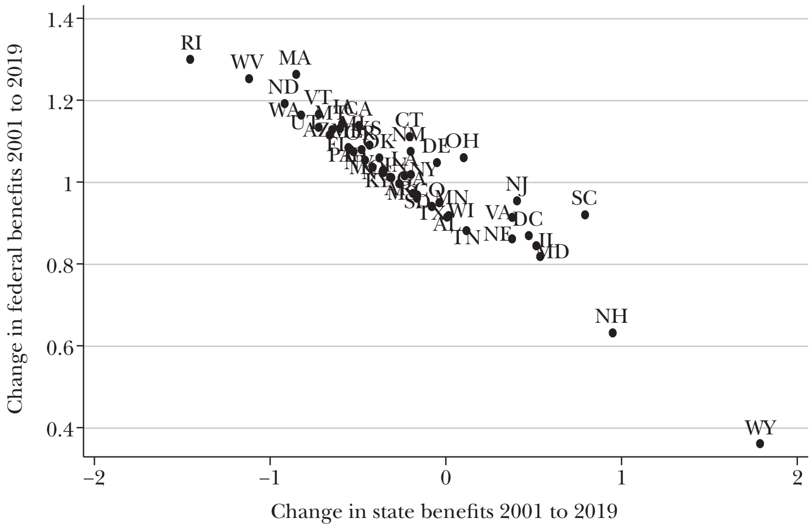
4) Resources available to married-parent families have expanded, whereas those for adults without dependents have remained thin. The original Aid to Families with Dependent Children program was only available to married-parent families in exceptional circumstances, and by the time of 1996 welfare reform there was substantial concern about marriage disincentives in the program. By contrast, the Earned Income Tax Credit and the Child Tax Credit offer benefits to both single and married parents, with the exact "marriage penalty" or "marriage bonus" determined by the earnings and tax filing choices of the individuals. Similarly, the Supplemental Nutrition Assistance Program is based on household composition regardless of marital status, and Medicaid expansions delinking health insurance from cash welfare also facilitated participation by married parents. Thus, the evolution of the safety net in recent decades has been particularly beneficial for low-income married couples with children. This is clearly illustrated by Figure 3a, which shows that total reported benefits increase the most between the two time periods for married-parent families (Figure 3a), as does public insurance coverage (Figure 3b). The only exception (as discussed earlier) is in the lowest earnings category, which experienced declines for married-parent and single-parent families.

The final panel in Figure 3 shows that individuals or married couples without children report extremely low safety net benefits relative to those with children. The maximum Earned Income Tax Credit available for a childless filer is under $700, less than 10 percent of the maximum for those with children, and the Supplemental Nutrition Assistance Program is available to low-income individuals without children with some restrictions. However, the recent Medicaid expansions have been particularly relevant for this group, as illustrated in the third panel of Figure 3b. Overall, cash support for low-income adults without dependents remains scant, and this group has largely been left behind in the evolution of the safety net over the past 25 years (Greenstein 2024).

5) State flexibility in the structure of state-directed benefits has grown, but federally-directed spending has expanded, leaving cross-state variability fairly constant. Since its inception, the Aid to Families with Dependent Children program had a greater level of state control over benefits and eligibility than other programs, leading to significant geographic variation across states in safety net benefits available to comparable families even before welfare reform. The passage of the welfare reform act in 1996 led to additional state flexibility with welfare payments, which could have exacerbated

Figure 5

Changes 2001–2019 in Simulated State and Federal Benefit Generosity



Source: Authors' creation.
Note: 2022 dollars. Federal benefits include federal EITC, refundable federal CTC, and SNAP. State benefits include state EITC and TANF. Simulated using a pooled consistent sample of single-parent, nondisabled, nonimmigrant families in the 2010–2019 CPS ASEC. Data points represent the average imputed benefits for this set of families as if they lived in a given state and year and they fully accessed benefits for which they are eligible given observed earnings and family structure. Simulation does not model work requirements or incomplete take-up.

differences in safety net generosity across states. Policy experimentation with state Earned Income Tax Credits also pushed towards increased variability across states in safety net generosity.

However, the expansion in the federally-directed Supplemental Nutrition Assistance Program and federal tax credit programs raised the benefit floor and counteracted the tendency towards rising variation across states. While Figure 2 illustrates that the increase in total safety net generosity was seen in almost all states, sometimes changing the relative ranking of states in terms of generosity, federally-directed increases were particularly important in states with the least generous policies.

This can be seen in Figure 5, which shows the change in the index of generosity in state-directed programs on the horizontal axis and the change in the index for federally-directed programs on the vertical axis. Federally-directed benefits grew everywhere (that is, all points are above zero on the vertical axis). Moreover, because reducing state-level welfare payments mechanically leads to eligibility for greater benefits in the Supplemental Nutrition Assistance Program, the states with the biggest gains in federal support were those with the most sizable Temporary Assistance for Needy Families declines, resulting in a generally negative relationship between federal and state changes. Because of the role of the federal government, the increased degree of variation seen in state-directed support has not carried over to the total cash and near-cash benefit package.

6) Shifts towards federally-directed spending have facilitated the ability of national safety net policy to respond in times of economic distress. By reaching more people and providing more benefits during recessionary periods, the safety net provides an insurance function and acts as an automatic stabilizer (Hoynes and Luttmer 2011). However, some programs fill this role more effectively than others. The Supplemental Nutrition Assistance Program provides benefits to non-earners, and thus is well-poised to respond to economic downturns, while programs like tax credits with earnings requirements may face challenges in periods of high unemployment (Bitler, Hoynes, and Kuka 2017; Bitler, Hoynes, and Schanzenbach 2020). Programs with high administrative burdens for enrollment such as Temporary Assistance for Needy Families are less able to reach new beneficiaries quickly in times of need.

At the same time, recent history suggests program rules are often adjusted in times of economic crisis, with expansions of eligibility and benefits and reductions in barriers to program entry. In particular, whereas states often face balanced budget rules that limit spending, the federal government does not face these same constraints and may be able to move quickly to expand federally-directed programs (Rueben, Randall, and Boddupalli 2018).

Shifts in program rules in recessionary times are evident in several periods since the 1996 welfare reform. For example, the 2001 recession coincided with expansions in refundable tax credits and the encouragement of state-level initiatives to make the Supplemental Nutrition Assistance Program more accessible. Such changes were even more pronounced in the Great Recession from 2007 to 2009. The American Recovery and Reinvestment Act of 2009 created an Emergency Contingency Fund to provide additional Temporary Assistance for Needy Families benefits to the states, and expanded SNAP benefit levels and waived SNAP work requirements. It also expanded the Earned Income Tax Credit for families with three or more children, and it increased the refundability threshold for the Child Tax Credit.

The safety net response to the COVID crisis was unprecedented, as shown by the spikes in spending in Figure 1.[6] In 2020, Supplemental Nutrition Assistance Program benefit levels were boosted by household emergency allotments, which remained in place until 2022 or early 2023 depending on the state. A temporary 2021 expansion of the Child Tax Credit substantially increased the level of the benefit and allowed non-earners to receive the full amount.[7] Part of the expanded CTC was also issued monthly. While the expanded federal CTC was only in effect

[6] This figure understates the response to the pandemic recession, as it only includes means-tested programs, and as such does not show the expansions of unemployment insurance and the Economic Impact Payments that made up the bulk of the early policy response (Bitler, Hoynes, and Schanzenbach 2020, 2023; Meyer, Han, and Sullivan 2024).

[7] Prior to the pandemic, the National Academies had convened a Committee on Building an Agenda to Reduce the Number of Children in Poverty by Half in 10 Years. Their report, issued in 2019, compared

for one tax year, more than a dozen states have added a state CTC in the past few years. Continuous coverage for Medicaid, as described above, increased Medicaid participation by 31 percent between 2020 and 2023 (Dague and Ukert 2023). Arguably, the federal safety net expansions in SNAP, Medicaid, and tax policy since 1996 facilitated the rapid and dramatic policy response during the COVID era.

Additional Distributional Implications

Effects of Changes by Race and Ethnicity

As noted above, the overall package of benefits available to low-income families has grown substantially over the three decades. Given long-standing racial differences in labor market opportunities and earnings, due in part to legacies of structural racism, Black families tend to be concentrated in lower parts of the earnings distribution, and therefore disproportionately benefit from the overall expansions in the safety net. In our analysis of simulated eligibility for single-parent, nonimmigrant families, the combined generosity of the cash and food safety net grew 15.7 percent for a typical Black family, 13.8 percent for a typical White family, 11.8 percent for a typical Hispanic family, and 11.5 percent for a typical Asian family between 2001 and 2019.

The distribution of safety net benefits between state- and federally-directed programs, and the recent shift towards federally-directed benefits, also have differential implications across race and ethnicity groups. As a starting point, Black families are more likely to live in states with less generous safety net benefits, while Asian families disproportionately live in states with higher levels of safety net generosity. Evidence suggests that state decisions about program rules are correlated with the racial composition of the population; for example, a higher share of Black residents is associated with a lower share of Temporary Assistance for Needy Families dollars spent on cash assistance (Hardy, Samudra, and Davis 2019) and with additional TANF funds spent on discouraging unmarried childbearing and promoting marriage (Parolin 2021). As a result, we might expect state-level differences in safety net generosity to translate into racial and ethnic differences in safety net generosity solely due to the states in which different groups reside.

However, because states in which Black families tend to live started with less generous Temporary Assistance for Needy Families benefits even before the welfare reform of 1996, there was less scope for further decline. In southern states with disproportionate Black populations, declines in TANF payments after the welfare reform were more than offset by increased eligibility benefits for Supplemental Nutrition Assistance Program benefits, as well as increases in the federal tax credits.

---

a number of policy options, and this helped inform the 2021 CTC expansion (National Academies of Sciences, Engineering, and Medicine 2019).

The increasing federalization of benefits did not eliminate safety net differences stemming from differences in residential location by race and ethnicity, but it did mean that racial gaps in safety net access did not increase over time. This finding aligns with prior research showing that expansions of federal programs reduce racial inequities. For example, Hardy, Hokayem, and Ziliak (2022) discuss the Earned Income Tax Credit and Buchmueller et al. (2016) focus on the expansion of Medicaid under the Patient Protection and Affordable Care Act of 2010.

Safety Net Policies and Immigrants

Most of the analysis presented here has been limited to safety net access for families comprised of US-born people. However, rules about access to safety net programs for immigrant populations vary by program and by state, and hinge on factors such as legal status and time in the United States.

Undocumented immigrants are ineligible for most major federal safety net programs, including Temporary Assistance for Needy Families, the Supplemental Nutrition Assistance Program, the Earned Income Tax Credit, the Child Tax Credit, and Medicaid.[8] There is some nuance to this statement, however, because families may receive reduced TANF and SNAP benefits on behalf of citizen children. In addition, undocumented federal tax filers can claim the CTC on behalf of citizen dependents, but cannot claim the EITC. Citizen children are also able to qualify for Medicaid coverage regardless of their parents' legal status, and prenatal care for the undocumented population is available in some states (Miller, Wherry, and Aldana 2024). Using state funds, a number of states provide health coverage to undocumented children, and California recently started offering Medicaid coverage to all income-eligible undocumented residents (Pillai, Pillai, and Artiga 2024).

The ability of newly arrived *legal* immigrants to access safety net benefits is also severely curtailed. The 1996 welfare reform contained provisions eliminating federal support for most legal residents for their first five years in the country. In addition, green card recipients eligible through a family-based immigration pathway must have a sponsor who files an affidavit of support; the income of the sponsor is typically considered in determining program eligibility. For the Supplemental Nutrition Assistance Program, children are exempted from the five-year waiting period, and many states have exercised the Medicaid option to waive the five-year ban for children and pregnant people (Pillai, Pillai, and Artiga 2024). Some states also offer "fill-in" programs using state funds to provide benefits for legal residents during the first five years in the country (East 2018).

Even when immigrant families are eligible for programs, take-up of safety net benefits for immigrants may be further constrained by "chilling effects." Undocumented parents may fear interacting with authorities on behalf of their citizen

[8] Other nutrition programs, notably the Women, Infants, and Children program and school-based programs, are available without regard to legal status. Undocumented immigrants may also be covered by Emergency Medicaid in the event they receive emergency health treatment for acute conditions.

children. Legal noncitizens may believe that participating in programs will affect their future pathway to citizenship. These chilling effects seem to be amplified during times of heightened enforcement as indicated by reduced participation even among those who are eligible (Watson 2014; Alsan and Yang 2024). Given that immigrants are less likely to identify as non-Hispanic White than the general population, policies restricting or suppressing immigrant access to the safety net also have implications for racial disparities (Bitler et al. 2021).

Conclusion

The safety net is important. A sizeable literature has documented its role in redistributing resources and insuring against income shocks (Hoynes and Luttmer 2011), in reducing material hardship (Schmidt, Shore-Sheppard, and Watson 2016; McKernan, Ratcliffe, and Braga 2021), and in improving child outcomes (Page 2024; Aizer, Hoynes, and Lleras-Muney 2022). Though the programs we study combined are less than one-tenth of the federal budget (and less than 4 percent excluding Medicaid), debates around their size and structure remain contentious.

The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 kept President Clinton's 1992 campaign promise: the welfare reform act did end welfare as we had known it. Since then, the safety net has both expanded and changed character—moving away from monthly cash transfers towards in-kind supports and refundable tax credits, providing more to the working poor and near-poor and less to families without earnings, and offering low-income married-parent families more support while doing little for adults without dependents. States enjoy more discretion than they had previously in the design of cash welfare programs, once a cornerstone of the safety net, but the relative importance of state-directed programs has shrunk over time. On net, because of significant expansions in federal programs, variability in the overall safety net generosity across states did not increase as those predicting a race to the bottom had feared. Significant cross-state inequities persist, however, with Black families disproportionately residing in the least generous states.

Despite expanded generosity and improved work incentives, limitations remain in the reshaped US social safety net. First, some groups have been left out by benefit expansions, limiting the ability of the safety net to perform its insurance and redistributive functions. The changes in the safety net described above have boosted incomes for many low-income children, but have reduced support for the most vulnerable children—those living in households with no workers. In addition, nonelderly, nondisabled individuals without children have been largely excluded from most forms of safety net support (Greenstein 2024).

In addition, the current safety net structure is centered on in-kind benefits (Medicaid, the Children's Health Insurance Program, and the Supplemental Nutrition Assistance Program) and on tax credits that are typically paid once per year. It provides little in consistent monthly cash transfers, yet many monthly needs cannot be met with in-kind benefits (Gennetian and Magnuson 2022). The adverse effects of this structure are felt most acutely by those who are less able to access cash income from other sources to finance nonfood purchases and those with limited access to credit. In recognition of these limitations, there is growing interest in the possibility of unconditional cash transfer programs or universal basic income programs (for discussion, see Shah and Gennetian 2024; Hoynes and Rothstein 2019).

Our conclusion is that the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 did not end welfare, but did end welfare as we knew it. The safety net provides much more support than in the past to low-income families overall, but it increasingly fails to meet the needs of adults without dependents and of families without earnings, who may have the greatest need. As traditional cash welfare continues to erode, the future of the support for low-income families will increasingly be negotiated in the sphere of in-kind benefit and tax policy.

■ *We are grateful to Michael Levere, Hamish Low, Jonathan Parker, Nina Pavcnik, Timothy Taylor, and Heidi Williams for helpful comments and to Janina Broker, Gabriela Campos, Ellie Chungath, Jillian da Costa Pinto, Gabriela Goodman, Simran Kalkat, Emma Phillips, Quinn Sanderson, Lily Smetzer, and Jonathon Zars for excellent research assistance.*

References

Aizer, Anna, Hoynes, Hilary, and Adriana Lleras-Muney. 2022. "Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children." *Journal of Economic Perspectives* 36 (2): 149–74.

Alsan, Marcella, and Crystal S. Yang. 2024. "Fear and the Safety Net: Evidence from Secure Communities." *Review of Economics and Statistics* 106 (6): 1427–41.

Azevedo-McCaffrey, Diana, and Tonanziht Aguas. 2024. "Continued Increases in TANF Benefits are Critical to Helping Families Meet Their Needs and Thrive." Center on Budget and Policy Priorities, May 29. https://www.cbpp.org/sites/default/files/5-29-24tanf.pdf.

Baicker, Katherine, Amy Finkelstein, Jae Song, and Sarah Taubman. 2014. "The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment." *American Economic Review* 104 (5): 322–28.

Bastian, Jacob. 2020. "The Rise of Working Mothers and the 1975 Earned Income Tax Credit." *American Economic Journal: Economic Policy* 12 (3): 44–75.

Bauer, Lauren, and Chloe N. East. 2023. *A Primer on SNAP Work Requirements.* Hamilton Project.

Beatty, Timothy K. M., and Charlotte J. Tuttle. 2015. "Expenditure Response to Increases in In-Kind Transfers: Evidence from the Supplemental Nutrition Assistance Program." *American Journal of Agricultural Economics* 97 (2): 390–404.

Ben-Shalom, Yonatan, Robert Moffitt, and John Karl Scholz. 2012. "An Assessment of the Effectiveness of Anti-Poverty Programs in the United States." In *The Oxford Handbook on the Economics of Poverty*, edited by Philip N. Jefferson, 709–49. Oxford: Oxford University Press.

Bitler, Marianne, Hilary Hoynes, and Elira Kuka. 2017. "Do In-Work Tax Credits Serve as a Safety Net?" *Journal of Human Resources* 52 (2): 319–50.

Bitler, Marianne, Lisa A. Gennetian, Christina Gibson-Davis, and Marcos A. Rangel. 2021. "Means-Tested Safety Net Programs and Hispanic Families: Evidence from Medicaid, SNAP, and WIC." *ANNALS of the American Academy of Political and Social Science* 696 (1): 274–305.

Bitler, Marianne, Hilary W. Hoynes, and Diane Whitmore Schanzenbach. 2020. "The Social Safety Net in the Wake of COVID-19." *Brookings Papers on Economic Activity* 51 (S): 119–58.

Bitler, Marianne P., Hilary W. Hoynes, and Diane Whitmore Schanzenbach, 2023. "Suffering, the Safety Net, and Disparities during COVID-19." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 9 (3): 32–59.

Blank, Rebecca M. 1997. "Policy Watch: The 1996 Welfare Reform." *Journal of Economic Perspectives* 11 (1): 169–77.

Bound, John, Sherri Ann Kossoudji, and Gema Ricart-Moes. 1998. "The Ending of General Assistance and SSI Disability Growth in Michigan." In Growth in Disability Benefits: Explanations and Policy Implications, edited by Kalman Rupp and David C. Stapleton, 223–48, 275–79. W. E. Upjohn Institute for Employment Research.

Buchmueller, Thomas C., Zachary M. Levinson, Helen G. Levy, and Barbara L. Wolfe. 2016. "Effect of the Affordable Care Act on Racial and Ethnic Disparities in Health Insurance Coverage." *American Journal of Public Health* 106 (8): 1416–21.

Buchmueller, Thomas, John C. Ham, and Lara D. Shore-Sheppard. 2016. "The Medicaid Program." In *Economics of Means-Tested Transfer Programs in the United States*, Vol. 1, edited by Robert A. Moffitt, 21–136. University of Chicago Press.

Burney, Shaheer, Rebecca Boehm, and Rigoberto Lopez. 2021. "The Impact of the ACA Medicaid Expansion on SNAP Participation." *Food Policy* 101: 102094.

Campbell, Colin, and S. Michael Gaddis. 2017. "'I Don't Agree with Giving Cash': A Survey Experiment Examining Support for Public Assistance." *Social Science Quarterly* 98 (5): 1352–73.

Cha, Paulette, and José J. Escarce. 2022. "The Affordable Care Act Medicaid Expansion: A Difference-in-Differences Study of Spillover Participation in SNAP." *PLoS ONE* 17 (5): e0267244.

Chatfield, Karen. 2024. "A 50-State Comparison of TANF Benefit Amounts." Factsheet, National Center for Children in Poverty, December 17. https://www.nccp.org/publication/a-50-state-comparison-of-tanf-amounts/.

Collinson, Robert, Ingrid Gould Ellen, and Jens Ludwig. 2016. "Low-Income Housing Policy." In *Economics of Means-Tested Transfer Programs in the United States*, Vol. 2, edited by Robert A. Moffitt, 59–126. University of Chicago Press.

Congressional Research Service. 2021. *Federal Spending on Benefits and Services for People with Low Income: FY2008–FY2020*. Congressional Research Service Report R46986.

Congressional Research Service. 2023. *School Meals and Other Child Nutrition Programs: Background and Funding*. Congressional Research Service Report R46234.

Congressional Research Service. 2024. *The Temporary Assistance for Needy Families (TANF) Block Grant.* Congressional Research Service Report IF10036.

Cunningham, Jamein, and Andrew Goodman-Bacon. 2025. "Changes in Family Structure and Welfare Participation since the 1960s: The Role of Legal Services." *American Economic Journal: Applied Economics* 17 (1): 369–401.

Currie, Janet, 2006. "The Take-up of Social Benefits." In *Poverty, the Distribution of Income, and Public Policy*, edited by Alan J. Auerbach, David Card, and John M. Quigley, 80–148. Russell Sage Foundation.

Currie, Janet, and Jonathan Gruber. 1996. "Health Insurance Eligibility, Utilization of Medical Care, and Child Health." *Quarterly Journal of Economics* 111 (2): 431–66.

Dague, Laura, and Benjamin Ukert. 2023. "Pandemic-Era Changes to Medicaid Enrollment and Funding: Implications for Future Policy and Research." *Journal of Policy Analysis and Management* 43 (4): 1229–59.

Duggan, Mark, Melissa S. Kearney, and Stephanie Rennane. 2016. "The Supplemental Security Income Program." In *Economics of Means-Tested Transfer Programs in the United States*, Vol. 2, edited by Robert A. Moffitt, 1–58. University of Chicago Press.

East, Chloe N. 2018. "Immigrants' Labor Supply Response to Food Stamp Access." *Labour Economics* 51: 202–26.

Edelman, Peter. 1997. "The Worst Thing Bill Clinton Has Done." *Atlantic,* March 1997. https://www.theatlantic.com/magazine/archive/1997/03/the-worst-thing-bill-clinton-has-done/376797/.

Falk, Gene. 2023. *Temporary Assistance for Needy Families (TANF) and Work Requirements.* Congressional

Research Service.

Falk, Gene. 2014. *Temporary Assistance for Needy Families (TANF): Eligibility and Benefit Amounts in State TANF Cash Assistance Programs.* Congressional Research Service.

Feenberg, Daniel Richard, and Elizabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis* 12 (1): 189–194.

Feenberg, Daniel Richard, and Elizabeth Coutts. 2024. "TAXSIM Version 35." https://taxsim.nber.org/taxsim35/ (accessed November 14, 2024).

Floyd, Ife, LaDonna Pavetti, Laura Meyer, Ali Safawi, Liz Schott, Evelyn Bellew, and Abigail Magnus. 2021. "TANF Policies Reflect Racist Legacy of Cash Assistance: Reimagined Program Should Center Black Mothers." Center on Budget and Policy Priorities, August 5. https://www.cbpp.org/research/income-security/tanf-policies-reflect-racist-legacy-of-cash-assistance.

Fox, Liana E., Misty L. Heggeness, José Pacas, and Kathryn Stevens, 2017. "Precision in Measurement: Using SNAP Administrative Records to Evaluate Poverty Measurement." Social, Economic, and Housing Statistics Division Working Paper 49.

Fuller, Sheridan. 2023. "New Deal, Same Compromise? Measuring the Effect of Racially Linked Welfare Policies on Program Participation and Children's Development." Unpublished.

Ganong, Peter, and Jeffrey B. Liebman. 2018. "The Decline, Rebound, and Further Rise in SNAP Enrollment: Disentangling Business Cycle Fluctuations and Policy Changes." *American Economic Journal: Economic Policy* 10 (4): 153–76.

Gennetian, Lisa A., and Katherine Magnuson. 2022. "Three Reasons Why Providing Cash to Families With Children Is a Sound Policy Investment." Center on Budget and Policy Priorities, May 11. https://www.cbpp.org/sites/default/files/5-11-22tanf.pdf.

Goodman-Bacon, Andrew, and Lucie Schmidt. 2020. "Federalizing Benefits: The Introduction of Supplemental Security Income and the Size of the Safety Net." *Journal of Public Economics* 185: 104174.

Greenstein, Robert. 2024. "The Austere US Safety Net for Poor, Non-elderly Adults Who Are Not Raising Children and Do Not Receive Disability Benefits." Brookings, April 17. https://www.brookings.edu/articles/the-austere-us-safety-net-for-poor-non-elderly-adults-who-are-not-raising-children-and-do-not-receive-disability-benefits/.

Hahn, Heather, David Kassabian, and Sheila Zedlewski. 2012. "TANF Work Requirements and State Strategies to Fulfill Them." Urban Institute, March. https://www.urban.org/sites/default/files/publication/25376/412563-TANF-Work-Requirements-and-State-Strategies-to-Fulfill-Them.PDF.

Halpern-Meekin, Sarah, Kathryn Edin, Laura Tach, and Jennifer Sykes. 2015. *It's Not Like I'm Poor: How Working Families Make Ends Meet in a Post-Welfare World.* 1st ed. University of California Press.

Han, Jeehoon, Bruce D. Meyer, and James X. Sullivan. 2022. "Real-Time Poverty, Material Well-Being, and the Child Tax Credit." *National Tax Journal* 75 (4): 817–46.

Hardy, Bradley, Charles Hokayem, and James P. Ziliak. 2022. "Income Inequality, Race, and the EITC." *National Tax Journal* 75 (1): 149–67.

Hardy, Bradley L., Rhucha Samudra, and Jourdan A. Davis. 2019. "Cash Assistance in America: The Role of Race, Politics, and Poverty." *Review of Black Political Economy* 46 (4): 306–24.

Hastings, Justine, and Jesse Shapiro. 2018. "How are SNAP Benefits Spent? Evidence from a Retail Panel." *American Economic Review* 108 (12): 3403–3540.

Herd, Pamela, and Donald P. Moynihan. 2018. *Administrative Burden: Policymaking by Other Means.* Russell Sage Foundation.

Hoynes, Hilary W., and Erzo F. P. Luttmer. 2011. "The Insurance Value of State Tax-and-Transfer Programs." *Journal of Public Economics* 95 (11–12): 1466–84.

Hoynes, Hilary W., and Ankur J. Patel. 2018. "Effective Policy for Reducing Poverty and Inequality? The Earned Income Tax Credit and the Distribution of Income." *Journal of Human Resources* 53 (4): 859–90.

Hoynes, Hilary W., and Jesse Rothstein. 2019. "Universal Basic Income in the US and Advanced Countries." NBER Working Paper 25538.

Hoynes, Hilary W., and Diane Whitmore Schanzenbach. 2009. "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program." *American Economic Journal: Applied Economics* 1 (4): 109–39.

Hoynes, Hilary, and Diane Whitmore Schanzenbach. 2016. "US Food and Nutrition Programs." In *Economics of Means-Tested Transfer Programs in the United States*, Vol. 1, edited by Robert A. Moffitt, 219–301. University of Chicago Press.

Hoynes, Hilary W., and Diane Whitmore Schanzenbach. 2018. "Safety Net Investments in Children."

NBER Working Paper 24594.

Jones, Maggie R., and James P. Ziliak. 2022. "The Antipoverty Impact of the EITC: New Estimates from Survey and Administrative Tax Records." *National Tax Journal* 75 (3): 451–79.

Katz, Michael B. 1989. *The Undeserving Poor: From the War on Poverty to the War on Welfare*. Pantheon.

Kearney, Melissa S., and Lesley J. Turner. 2013. "Giving Secondary Earners a Tax Break: A Proposal to Help Low- and Middle-Income Families." Hamilton Project Discussion Paper 2013-07.

Ko, Wonsik, and Robert A. Moffitt. 2022. "Take-Up of Social Benefits." NBER Working Paper 30148.

Kubik, Jeffrey D. 2003. "Fiscal Federalism and Welfare Policy: The Role of States in the Growth of Child SSI." *National Tax Journal* 56 (1): 61–79.

Matsudaira, Jordan D., and Rebecca M. Blank. 2014. "The Impact of Earnings Disregards on the Behavior of Low-Income Families." *Journal of Policy Analysis and Management* 33 (1): 7–35.

McKernan Signe-Mary, Caroline Ratcliffe, and Breno Braga. 2021. "The Effect of the US Safety Net on Material Hardship over Two Decades." *Journal of Public Economics* 197: 104403.

Meyer, Bruce D., Jeehoon Han, and James X. Sullivan. 2024. "Poverty, Hardship, and Government Transfers." NBER Working Paper 33052.

Meyer, Bruce D., and Nikolas Mittag, 2019. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net." *American Economic Journal: Applied Economics* 11 (2): 176–204.

Meyer, Bruce D., Nikolas Mittag, and Robert M. Goerge, 2022. "Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation." *Journal of Human Resources* 57 (5): 1605–44.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29 (4): 199–226.

Miller, Sarah, Laura Wherry, and Gloria Aldana. 2022. "Covering Undocumented Immigrants: The Effects of a Large-Scale Prenatal Care Intervention." NBER Working Paper 30299.

Mitchell, Alison. 1996. "Two Clinton Aides Resign to Protest New Welfare Bill." *New York Times*, September 12. https://www.nytimes.com/1996/09/12/us/two-clinton-aides-resign-to-protest-new-welfare-law.html.

Moffitt, Robert A. 2015. "The Deserving Poor, the Family, and the US Welfare System." *Demography* 52 (3): 729–49.

Moffitt, Robert A. 2016. "Introduction." In *Economics of Means-Tested Transfer Programs in the United States*, Vol. 1, edited by Robert A. Moffitt, 1–20. University of Chicago Press.

Moffitt, Robert, and Anuradha Rangarajan. 1991. "The Work Incentives of AFDC Tax Rates: Reconciling Different Estimates." *Journal of Human Resources* 26 (1): 165–79.

Nichols, Austin, and Jesse Rothstein. 2016. "The Earned Income Tax Credit." In *Economics of Means-Tested Transfer Programs in the United States*, Vol. 1, edited by Robert A. Moffitt, 137–218. University of Chicago Press.

National Bureau of Economic Research (NBER). 2025. "Monthly Recession Indicators [USRECM]." Retrieved from FRED, Federal Reserve Bank of St. Louis. http://fred.stlousifed.org/series/USRECM (accessed January 7, 2025).

National Academies of Sciences, Engineering, and Medicine. 2019. *A Roadmap to Reducing Child Poverty*. Edited by Greg Duncan and Suzanne Le Menestrel. National Academies Press.

National Research Council. 1995. *Measuring Poverty: A New Approach*. Edited by Constance F. Citro and Robert T. Michaels. National Academies Press.

Page, Marianne E. 2024. "New Advances on an Old Question: Does Money Matter for Children's Outcomes?" *Journal of Economic Literature* 62 (3): 891–947.

Parolin, Zachary. 2021."Temporary Assistance for Needy Families and the Black–White Child Poverty Gap in the United States." *Socio-Economic Review* 19 (3): 1005–35.

Pavetti, LaDonna, Ali Safawi, and Danilo Trisi. 2021. "TANF at 25: A Weaker Cash Safety Net Reaching Fewer Families and Doing Less to Lift Families Out of Deep Poverty." *National Tax Journal* 74 (3): 763–89.

Perl, Libby. 2018. *LIHEAP: Program and Funding*. Congressional Research Service

Pillai, Akash, Drishti Pillai, and Samantha Artiga. 2024. "State Health Coverage for Immigrants and Implications for Health Coverage and Care." *KFF*, May 1. https://www.kff.org/racial-equity-and-health-policy/issue-brief/state-health-coverage-for-immigrants-and-implications-for-health-coverage-and-care/.

Romich, Jennifer L., and Thomas Weisner. 2000. "How Families View and Use the EITC: Advance

Payment versus Lump Sum Delivery." *National Tax Journal* 53 (4.2): 1245–65.

Rueben, Kim S., Megan Randall, and Aravind Boddupalli. 2018. *Budget Processes and the Great Recession.* Urban Institute.

Schmidt, Lucie, and Purvi Sevak. 2004. "AFDC, SSI, and Welfare Reform Aggressiveness: Caseload Reductions versus Caseload Shifting." *Journal of Human Resources* 39 (3): 792–812.

Schmidt, Lucie, Lara Shore-Sheppard, and Tara Watson. 2016. "The Effect of Safety Net Programs on Food Insecurity." *Journal of Human Resources* 51 (3): 589–614.

Schmidt, Lucie, Lara Shore-Sheppard and Tara Watson. 2021a. "The Impact of Expanding Public Health Insurance on Safety Net Program Participation: Evidence from the ACA Medicaid Expansion." NBER Working Paper 26504.

Schmidt, Lucie, Lara Shore-Sheppard and Tara Watson. 2021b. "Safety Net Program Interactions and Impacts on Low-Income Families." *NBER Reporter* 4: 10–15.

Schmidt, Lucie, Lara Shore-Sheppard, and Tara Watson. 2023. "The Effect of Safety Net Generosity on Maternal Mental Health and Risky Health Behaviors." *Journal of Policy Analysis and Management* 42 (3): 706–36.

Schmidt, Lucie, Lara Shore-Sheppard, and Tara Watson. 2025. Data and Code for: "Did Welfare Reform End the Safety Net as We Knew It? The Record since 1996." Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E214021V1.

Shah, Hema, and Lisa A. Gennetian. 2024. "Unconditional Cash Transfers for Families with Children in the US: A Scoping Review." *Review of Economics of the Household* 22 (2): 415–50.

Shivastava, Aditi, and Maria Manansala. 2024. "To Strengthen Economic Security and Advance Equity, States Should Invest More TANF Dollars in Basic Assistance." Center on Budget and Policy Priorities, September 23. https://www.cbpp.org/research/income-security/to-strengthen-economic-security-and-advance-equity-states-should-invest-0.

Soss, Joe, and Sanford F. Schram. 2006. "Welfare Reform as a Failed Political Strategy: Evidence and Explanations for the Stability of Public Opinion." *Focus* 24 (3): 17–23.

Tax Policy Center. 2024a. "Earned Income Tax Credit Parameters 1975 to 2024." https://taxpolicycenter.org/statistics/eitc-parameters (accessed January 12, 2025).

Tax Policy Center. 2024b. "How do state earned income tax credits work?" *The Tax Policy Briefing Book.* https://taxpolicycenter.org/briefing-book/how-do-state-earned-income-tax-credits-work (accessed January 12, 2025).

US Bureau of Economic Analysis (BEA). 2025. "Population [B230RC0A052NBEA]," retrieved from FRED, Federal Reserve Bank of St. Louis https://fred.stlouisfed.org/series/B230RC0A052NBEA (accessed January 7, 2025).

US Bureau of Labor Statistics (BLS). 2025. "Consumer Price Index", retrieved from Federal Reserve Bank of Minneapolis at https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913- (accessed January 7, 2025).

US Census Bureau. 2001–2003 and 2017–2019. *American Community Survey (ACS): Public Use Microdata Sample (PUMS), 2001–2003 and 2017–2019.* Washington, DC: US Department of Commerce; distributed by National Bureau of Economic Research Cambridge, MA. https://www.nber.org/research/data/current-population-survey-cps-supplements-annual-demographic-file (accessed February 17, 2017).

US Department of Agriculture. 2024. "Fiscal Year 2024 Cost-of-Living Adjustments." https://fns-prod.azureedge.us/sites/default/files/resource-files/FY24-SNAP-COLA-Memo.pdf (accessed January 10, 2025).

US Department of Agriculture. 2025. "Supplemental Nutrition Assistance Program—Key Statistics and Research." https://www.ers.usda.gov/topics/food-nutrition-assistance/supplemental-nutrition-assistance-program-snap/key-statistics-and-research (accessed January 12, 2025).

US General Accounting Office. 1994. *Families on Welfare: Sharp Rise in Never-Married Women Reflects Societal Trend.* General Accounting Office.

US Social Security Administration. 2005. *Annual Statistical Supplement to the Social Security Bulletin, 2005.* Washington, DC: Social Security Administration.

Watson, Tara. 2014. "Inside the Refrigerator: Immigration Enforcement and Chilling Effects in Medicaid Participation." *American Economic Journal: Economic Policy* 6 (3): 313–38.

Ziliak, James P. 2016. "Temporary Assistance for Needy Families." In *Economics of Means-Tested Transfer Programs in the United States,* Vol. 1, edited by Robert A. Moffitt. 303–93. University of Chicago Press.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 129–150

Administrative Burdens in the Social Safety Net

Pamela Herd and Donald Moynihan

By the age of five, Abel Sewell had survived cancer and was receiving monthly blood tests to ensure his leukemia had not returned. The tests were covered by TennCare, Tennessee's Medicaid program, until, at a regular doctor's visit, his mother discovered that the coverage had lapsed. Abel's mother spent months fighting to get their coverage restored, ultimately taking out a second mortgage on their home to manage the mounting health care debt that came from being uninsured. The family very much wanted health services, and were willing to endure significant hardships to get it.

The Sewells were not an anomaly. Between 2016 and 2019, almost 250,000 children in Tennessee lost coverage. What happened? The Sewells, like many others, said they never received the TennCare renewal forms. Even those who did receive the renewal packets often struggled to complete the 47 pages (Kelman and Reicher 2019). Failure to return forms accounted for 67 percent of those who lost coverage, and it seems likely that many of them did not receive the forms due to outdated mailing addresses. Late or incomplete forms also resulted in a loss of coverage (Arbogast, Chorniy, and Currie 2022). How should policy researchers and policymakers understand what happened to Abel Sewell and others like him, instances in which a series of bureaucratic obstacles impede access to the very policies intended to help them?

■ *Pamela Herd is Carol Kakalec Kohn Professor of Social Policy and Donald Moynihan is the Ira and Nicki Harris Family Professor of Public Policy, both at the Ford School of Public Policy, University of Michigan, Ann Arbor, Michigan. Their email addresses are pherd@umich.edu and dmoyn@umich.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20231394.

Table 1
Categories of Administrative Burden

Learning costs	Time and effort expended to learn about the program or service, ascertaining eligibility status, the nature of benefits, conditions that must be satisfied, and how to gain access.
Compliance costs	Provision of information and documentation to demonstrate standing; financial costs to access services (such as fees, legal representation, travel costs); avoiding or responding to discretionary demands made by administrators; the amount of time spent on these processes.
Psychological costs	Stigma arising from applying for and participating in an unpopular program; loss of autonomy that comes from intrusive administrative supervision; frustration at dealing with learning and compliance costs, unjust or unnecessary procedures; stresses that arise from uncertainty about whether a citizen can negotiate processes and compliance costs; fear about the coercive face of state power.

Source: Adapted from Herd and Moynihan (2018).

There is growing attention to the pervasive problem of safety net benefits left unused by eligible individuals (National Academies of Sciences, Engineering, and Medicine 2023). As Table 2 shows, large fractions of those eligible for key social welfare policies do not actually receive those benefits. In some cases, this might reflect preferences of the client to forgo benefits. A $10 monthly food stamp benefit might not be worth the hassle of applying for and maintaining access to that benefit. But in other cases, especially when client needs are intense, the answer lies elsewhere: perhaps clients never apply for services they are unaware of, mistakenly believe they are ineligible for them, or cannot figure out how to get started. Perhaps they apply, but do not make it through hurdles in the application process: mandatory interviews, demands for documentation, or lengthy and confusing paperwork. They may also lose eligibility for the program when a renewal deadline arises. Even those that access benefits may struggle to use them, due to, for example, the complicated rules regarding what you can buy in food stamp programs, what Barnes (2021) calls redemption costs. At every step, administrative processes might be experienced as stigmatizing, disempowering, frustrating, and stressful.

In this article, we group together these issues under the heading of "administrative burden," which we define as the experience of policy implementation as onerous to the individual. The types of administrative burden costs can be cataloged as learning, compliance, and psychological costs, as shown in Table 1. These frictions are interrelated, and generate similar effects, across a variety of safety net programs.

Why do we need an administrative burden approach to better understand the safety net? As scholars working in interdisciplinary spaces, we have observed researchers from an array of disciplines and specific policy areas offer insights about the origins, nature, and impact of frictions in the operation of public services, but with relatively little communication with one another. The administrative burden approach is intended to address that problem. It does not rely on a single model, but a set of nested concepts and assumptions about empirical relationships that can be tested. The idea of administrative burden belongs to no single disciplinary field and is relatively recent in origin, and so our summary here is provisional. It offers a framework for those interested in how the public interacts with the state, the resulting consequences, and whether it is possible to restructure those interactions to produce better outcomes.

The article proceeds as follows. We review three core claims of the administrative burden approach. We next compare administrative burdens to other approaches to frictions. We then identify the emergence of a public policy infrastructure, via presidential executive orders and statutory guidance, to address these burdens. Finally, we consider some implications of the administrative burden research for the study of the safety net. While we focus on burdens in the American safety net, people experience burdens across a wide array of government services, and we conclude by discussing their relevance in other countries.

Basic Claims about Administrative Burden

Research on administrative burden is based on three foundational claims: (1) even small administrative burdens can have large effects; (2) administrative burdens can reinforce and exacerbate inequality; and (3) administrative burdens can be policymaking by other means. While we focus on the cost of burdens, it is important to note that burdens can be justified by other public service goals, such as minimizing fraud. Decisions on policy design and implementation should, however, be informed by evidence of the negative and positive effects of bureaucratic practices that generate burden, as well as a search for less onerous alternatives that satisfy core policy goals.

Small Burdens, Big Effects

Asking someone to complete a seemingly simple administrative process can prove a significant impediment, even when it is rational for individuals to complete that process. A public health insurance system that required individuals to pick and enroll in a plan—after having been deemed eligible for free coverage—led to a 33 percent drop in coverage compared to a system of automatic enrollment (Shepard and Wagner 2022). Simply removing Medicaid recipients' renewal requirements during the Covid-19 public health emergency led to a greater increases in health insurance coverage than the Patient Protection and Affordable Care Act of 2010, the most visible and contentious extension of public health insurance in decades (Dague and Ukert 2024). Indeed, when the public health emergency ended, nearly one-third of Medicaid beneficiaries lost their coverage once they were put through renewal processes again (Tolbert and Corallo 2024). It was not simply that people were no longer eligible. Over 70 percent of those disenrolled were for procedural reasons, such as missing paperwork (Tolbert and Corallo 2024). States

Table 2
Take-Up Rates of Large Safety Net Programs

Program	*Take-up rate*
Supplemental Nutrition Assistance Program	82%
Medicaid	50% for adults, 65% for children
Temporary Assistance to Needy Families	28%
Earned Income Tax Credit	77%

Source: National Academies of Science, Engineering, and Medicine (2023, pp. 128–29).

vary widely in these procedural denials, from a low of 23 percent in Maine to over 90 percent in states like Nevada, Utah, and New Mexico, which is consistent with burdensome processes, rather than eligibility rule themselves, as a key driver (Guth et al. 2023).

Even discrete steps within larger administrative processes can have large effects. In a study of the Supplemental Nutrition Assistance Program, commonly known as "food stamps," Homonoff and Somerville (2021) focused on the requirement to complete an interview with a caseworker as part of the annual process to ensure clients were still eligible. They found that clients who had a smaller time window to complete the interview, compared to those who had a larger window, were 22 percent less likely to get through the recertification process and continue to receive benefits. Conversely, efforts to reduce such frictions can have large impacts. An experiment that used administrative data to pre-fill forms for student loan recipients who were eligible to receive reduction in their loan payments increased take-up from 24 percent among those not receiving pre-filled forms to over 60 percent among those receiving pre-filled forms (Mueller and Yannelis 2022).

Burdens Can Reinforce Inequality

Speaking broadly, programs that target benefits to low-income, racially marginalized, and disabled populations tend to shift more of the administrative burdens for proving eligibility to individuals than do programs that confer benefits to higher income, white, and able-bodied individuals (Herd et al. 2023; National Academies of Sciences, Engineering, and Medicine 2023, pp. 132, 134–135; Herd and Johnson 2024; Ray, Herd, and Moynihan 2023). For example, Social Security retirement benefits have low administrative burdens, with the government collecting lifetime earnings data and automating most of the processes that determine eligibility and benefit size (Herd and Moynihan 2018). Nearly 100 percent of those eligible receive the benefit. In contrast, the Supplemental Security Income program, which benefits low-income, blind, disabled, and elderly individuals, is rife with complicated paperwork and documentation for participants to manage, with take-up rates around 60 percent for low-income older adults (McGarry and Schoeni 2015). In another example, the tax subsidies that individuals receive for employer-based health insurance coverage are nearly equivalent to Medicaid spending for low-income Americans, but while those receiving employer-based health insurance face no burdens to access that subsidy, the burdens are extraordinarily high for those accessing Medicaid (CBO 2024).

Beyond the greater presence of burdens in means-tested programs, groups may be reluctant to engage with administrative processes because of concerns about discrimination or other negative outcomes. For example, during the first Trump administration, when anti-immigration statements accelerated rapidly and general burdens in Medicaid increased (Barofsky et al. 2020), Hispanic families became less likely than non-Hispanic families to access health insurance, along with families that included a noncitizen (Arbogast, Chorney, and Currie 2022).

Those funneled into programs with higher administrative burdens also appear to be, on average, less likely to navigate those burdens successfully to access benefits; those who might most need help may also struggle the most to overcome burdens they experience in seeking the help (Christensen et al. 2020). A National Academies of Sciences, Engineering, and Medicine report (2023, p. 134) concludes, "Because these safety-net-programs are aimed at low-income families, the problem of cognitive barriers is particularly relevant: lower-than-average levels of education and literacy are common for this population, yet the administrative complexity of applying for the programs and continuing to receive benefits is significant. Low-income individuals and families often cope with significant daily life challenges, so limited attention is a significant factor for this population." Skills such as education may matter because they capture "administrative capital," which refers to the skills and knowledge an individual has to negotiate administrative spaces (Masood and Nisar 2021). Poor health (including mental health) and disability make it harder to navigate administrative requirements (Bell et al. 2023). For example, using a natural experiment based on closure of Social Security field offices, Deshpande and Li (2019) find that such closures discourage applications from clients with lower education, earnings, and with moderately severe disabilities. In the study mentioned earlier about interview completion to maintain eligibility for the Supplemental Nutrition Program, Homonoff and Somerville (2021) find that the lowest income individuals are the least likely to complete such an interview.

Burdens as Form of Policymaking

Administrative burdens generate policy effects: sometimes intended, sometimes not. In either case, the opaque nature of state actions that give rise to administrative burdens—which often depend upon complicated or obscure administrative processes—can make them especially attractive to policymakers. For example, anti-abortion legislators used a series of burdens, on both individuals and clinics, to restrict, in practice, access to legal abortions (Herd and Moynihan 2018, 2024). Burdens can also serve to ration services for policymakers seeking to minimize spending. Conversely, state actions that give rise to burdens may fail to achieve their stated purpose. As we discuss in greater detail below, the administrative burden involved in work requirements is often justified as a means to increase employment, but it often fails to achieve this goal while making it harder for eligible clients to receive benefits and services.

Comparing Administrative Burden to Other Approaches to Frictions

There are other ways of thinking about frictions (for a review, see Madsen, Mikkelsen, and Moynihan 2022). We review two alternative approaches: "ordeal mechanisms" and "sludge." Both are rooted in economic theory, but, as we will describe, they come to different conclusions regarding the relative value of administrative burden in social welfare programs.

Ordeal Mechanisms

Ordeal mechanisms originated in the economics literature focused on social policy. The original paper that coined the term defined these ordeals as the imposition of "deadweight costs" on individuals applying for public programs (Nichols and Zeckhauser 1982, p. 372). The appeal of ordeals was that they could serve as another way, beyond eligibility requirements, to target benefits more efficiently (Heinrich et al. 2022).

The ordeal mechanism approach assumed that individuals encountering burdens behave rationally. Those who do not truly need a service will be put off by the hassles and will exit out of burdensome processes. Burdens become the price, paid in time or other hassles, for obtaining a benefit, and needier individuals are willing (and able) to pay this price. In sum, "[t]he demeaning qualification tests and tedious administrative procedures involved in many [social welfare] programs may serve such a sorting function" (Nichols and Zeckhauser 1982, p. 372).

The administrative burden approach, in contrast, casts doubt on the value of burdens as a way to target benefits more effectively. Instead, it assumes that the ability (or inability) to overcome burdens will be driven by factors other than individual preferences and the relative value of time. For example, peoples' ability to navigate complicated administrative requirements may be affected by things such as educational attainment, financial resources, and experiences of poor health or scarcity. Those with fewer resources may have less bandwidth to engage with or complete forbidding or complex administrative processes, even if they deeply value or need the benefits that might follow (Christensen et al. 2020; Mullainathan and Shafir 2013).

Of course, whether or not ordeals improve targeting, or increase inequality, is an empirical question. But recent evidence challenges whether burdens are efficient at targeting. For example, time diary studies reveal that people with poorer health spend more time on administrative tasks and report more negative well-being because of their experiences with burdens (Martin, Delaney, and Doyle 2023). There is also evidence that burdens loom larger for public-service clients who report poorer mental and physical health, increasing their tendency to withdraw from programs (Bell et al. 2023; Deshpande and Li 2019).

Apart from the empirical effects of ordeals, any effort to use administrative hassles as a targeting tool raises a particular set of normative concerns. Policymakers design programs with specific eligibility requirements. Adding administrative requirements to limit access to individuals who meet those legal requirements raises questions about fairness in a democratic system. It is one thing to use an ordeal mechanism with the idea that it will discourage millionaires from consuming resources also used by lower-income clients; it is another to discourage a group of lower-income clients from claiming benefits *to which they are legally entitled* and from which they may benefit. A growing body of evidence that administrative burdens weigh most heavily on those with lower levels of human capital and health raises fundamental questions about how ordeals worsen equity in the delivery of public services.

Sludge

Sludge emerged from the growing field of behavioral economics, which has increasingly shown that people do not always make optimal decisions, due to a mixture of cognitive and psychological biases, to maximize their overall welfare (Tversky and Kahneman 1974; Thaler and Sunstein 2009). This approach challenges the assumption behind the ordeals approach, which is that people who need and want benefits will actually navigate burdensome processes.

Specifically, sludge refers to "[e]xcessive or unjustified frictions that make it difficult for consumers, employees, employers, students, patients, clients, small businesses, and many others to get what they want or to do as they wish" (Sunstein 2022, p. 656). The name "sludge" consciously echoes the "nudge" focus of behavioral economics, as in Thaler and Sunstein's (2009) eponymous book. The right "choice architecture," or a nudge, can help us overcome common human biases, like our tendency towards inertia or our cognitive limitations, so that we can make better choices. Sludge, present in both the public and private sector, instead exacerbates these human biases and limitations. Sunstein (2021, 2022) calls for regular "sludge audits" to address such problems, reflecting the point that governments typically do not revisit the need and value of rules and processes that have accumulated over time.

There is much in common between the administrative burden and sludge perspectives. Indeed, the burden approach draws from the insights of Thaler and Sunstein (2009), and also pioneering work by Moffitt (1983) on stigma and Currie (2006) on the reasons why individuals do not take up benefits. But there are some distinctions between the sludge and administrative burden framework. First, the administrative burden literature is more attentive to the political economy of frictions, including the political origins and feedback effects of state actions that beget burdens, such as how burdens can be "policymaking by other means" (Baekgaard, Halling, and Moynihan 2024; Baekgaard, Moynihan, and Thomsen 2021; Herd and Johnson 2024; Herd and Moynihan 2018). Second, the administrative burden work is more centered on the relationship between burdens and inequality, and also how inequality can instigate and reinforce burdens. For example, burdensome requirements are more prevalent in programs where beneficiaries are largely poor and racially marginalized women than in programs that benefit wealthier white men (Ray, Herd, and Moynihan 2022; Herd and Moynihan 2024).

Ultimately, the idea that frictions matter is relatively intuitive. Researchers working on the topic can draw from multiple approaches, and their choices may reflect disciplinary training or other factors (as discussed in the review of frictions in Madsen, Mikkelsen, and Moynihan 2022). Ordeal mechanisms suggest that using hassles to ration safety net benefits may significantly counterbalance any costs, making this an attractive approach for policymakers who want to limit spending. Conversely, a sludge approach suggests that simplifying safety net administrative processes to reduce costs of time and learning is worthwhile. But research on administrative burden points to the need for significant structural and organizational changes, as well as identifying when the roots of that administrative burden lie in political and value-based debates. Moreover, directing attention to the specific types of costs people encounter in safety net programs provides a framework for clarifying what constitutes the burden, as well as how to identify and design interventions to address it.

A Public Policy Infrastructure for Reduction of Administrative Burden

Attention to administrative burden has been growing, not only in research, but also in broader policy and practice. The administrative burden approach invites policymakers to reconsider existing policy designs, such as overly complicated eligibility criteria for programs, by considering the implementation of these policies from the perspective of users. But while we can reduce (or add) burdens via legislation, federal and state executive agencies can also significantly affect these burdens by changing institutional and organizational practices.

In recent years, an administrative burden policy framework has emerged at the federal government level, relying on executive orders, legislation, and agency guidance to signal the goal of burden reduction, as well as to provide agencies with the authority to enact change. Among the first executive orders that President Joseph Biden (2021a) signed, which was primarily focused on racial equity, tied reduction of administrative burden to inequality. Later that year, Biden (2021b) signed a second executive order that was even more explicit, stating, "Agencies must work with the Congress; the private sector and nonprofit organizations; State, local, Tribal, and territorial governments; and other partners to design experiences with the Federal Government that effectively reduce administrative burdens, simplify both public-facing and internal processes to improve efficiency, and empower the Federal workforce to solve problems." Attention to burden reduction was also incorporated into the Government Service Delivery Improvement Act signed by President Biden in early 2025.

While there have been prior efforts to improve customer service in government, the emerging policy infrastructure emphasis is different in three important ways. First, the connection between customer experience and inequality is new. Second, current efforts draw more directly from social science to inform its approach, rather than relying on business principles (US Office of Management and Budget 2021). Third, the current approach is more deeply embedded into existing administrative processes, and therefore less likely to fade away with a new president. Circular A-11, the budget guidance that every agency must follow, was also revised to direct agencies to report on burden reduction efforts (US Office of Management and Budget 2022). Guidance for the Paperwork Reduction Act, overseen by the Office of Information and Regulatory Affairs, was revised to encourage agencies to do more to identify learning, compliance, and psychological costs, and to reduce administrative burdens (US Office of Information and Regulatory Affairs 2022). The 21st Century Integrated Digital Experience Act of 2018 has been used to push agencies to create accessible and inclusive designs for all abilities. The Biden administration also drew on the Plain Writing Act of 2010 to push agencies to communicate clearly and concisely, as well as to encourage comments from the public on areas for improvement.

Bureaucracies have traditionally had few incentives to understand how they impose costs on members of the public. Creating leadership momentum, a policy infrastructure, and a set of tools to allow them to do so can help address this problem. For example, the Department of Homeland Security successfully achieved a goal of reducing 20 million hours of citizen time spent interacting with the agency (US Executive Office of the President 2023). The policy infrastructure of identifying administrative burdens, so that they can be reduced, has been created quickly. How it will evolve remains to be seen. In his first term in office, President Trump (2018) signed an executive order that called for more work requirements to be placed on safety net programs. Political disagreement about the size and value of safety net policies, and the communities they serve, may therefore lead them to take different approaches to supporting or opposing administrative burdens (Herd and Moynihan 2018). But for federal officials interested in burden reduction, a coherent policy framework now exists, one that state and local governments can choose to follow.

Implications for the Safety Net

Giving greater attention to administrative burdens can change how the policy design of safety net programs is understood, encouraging systems that will not become weighed down by unnecessary requirements. Similarly, attention to administrative burdens points to a problem with federalism, in which each level of government adds opportunities for new frictions, as well as problems with privatized and fragmented service delivery mechanisms that create confusion rather than generate competitive behaviors that benefit clients (Goldstein et al. 2023; Herd et al. 2023). The policy design lessons that follow also offer a useful agenda for researchers interested in evidence on how burdens matter and how to reduce them.

Avoid Predictably Ineffective State Actions

Not all administrative burdens are equal. Some have large negative effects on access, while providing few benefits. In particular, policy design motivated by the actions of a small population or by unusual situations, but which create frictions for a large population of eligible beneficiaries, will predictably lead to large costs with limited positive effects. Such state actions often reflect concerns about legitimate political values, such as minimizing fraud. But policymaking is, or should be, about both facts and values, and the research can bring evidence to the table about tradeoffs. The question "are we going to tolerate any fraud?" is very different from the question of "how much fraud are we willing to tolerate given impacts on access?" The latter question implies a willingness to consider tradeoffs, which in turn requires evidence. Private companies, such as credit card companies or grocery stores, make these tradeoffs all the time, accepting some measure of theft or fraud rather than imposing a set of requirements that make services impossible to access. Some examples clarify this point in the context of safety net programs.

Asset tests, or restricting eligibility to those with low (or no) assets (such as a home or savings), are an example of a predictably ineffective state action driven by fraud concerns. The target population of asset tests in poverty-based programs is presumably low-income individuals who are seeking safety net benefits but have significant wealth. In practice, this group is mostly confined to a relatively small group of older adults, but imposing asset tests on all generates significant costs (Ratcliffe et al. 2016). States that removed asset tests for the Supplemental Nutrition Assistance Program saw a 20 percent reduction in administrative costs and no increase in fraud, while increasing participation of those previously eligible by streamlining administrative processes (Lin 2023). In this setting, the administrative burden of asset tests raised program costs and offered little value in reducing fraud.

Another value-based concern is that those receiving benefits should be employed if they are able. Key parts of the existing social safety net, including Temporary Assistance to Needy Families (TANF), the Earned Income Tax Credit, and the Child Tax Credit are already tied to work. Policymakers disagree about expanding work requirements for food stamps (SNAP) and introducing new work requirements for health insurance programs like Medicaid (Herd and Moynihan 2021; Haeder and Moynihan 2023).

Work requirements assume a population of individuals who could be employed, but choose to rely on welfare benefits instead. Moreover, they assume a population that is not already subject to work requirements under other welfare programs. In practice, this population is very small.

Additional work requirements do little to spur additional labor force participation. For example, just 6 percent of those on Medicaid are nonemployed and nonelderly clients who would not qualify for an exemption from work requirements, and one-third of those 6 percent are retired (Guth et al. 2023). There is little evidence that work requirements in the Supplemental Nutrition Assistance Program have substantially increased labor force participation. However, such requirements have caused vulnerable groups, such as the homeless and those with chronic health conditions, to lose benefit access (Cook and East 2024; Cuffey, Beatty, and Mykerezi 2022; Gray et al. 2023; Han 2022; Ndumele et al. 2024). For this reason, some safety net advocates characterize work requirements as "work reporting requirements," to communicate that the consequential aspect of the regulation is the paperwork burdens, not on actual labor force participation.

When Arkansas adopted Medicaid work requirements, 95 percent of those targeted by the program were already employed or should have been exempted due to disability, but the work requirements in the state reduced Medicaid enrollment by 12 percentage points (Sommers et al. 2019). About one-third of those who lost coverage, who were disproportionately those with lower education, were simply unaware of the new requirements. The work requirements did not increase labor force participation. However, half of those who lost Medicaid coverage reported serious problems paying off medical debt, 56 percent delayed care due to cost, and 64 percent delayed taking medications because of cost (Sommers et al. 2020).

Policymakers should be able to gather information to understand which administrative burdens are having disproportionately negative effects. For example, the Biden administration altered its approach to requiring annual income verification processes for a student loan forgiveness program for people with disabilities, pointing to evidence that this requirement led to half of participants losing benefits, even though fewer than one in ten actually become ineligible because of income changes (US Executive Office of the President 2023). The administrative burdens of state actions like asset tests or work requirements—and the effects of these burdens on program participation—are not unanticipated (Cook and East 2024). If policymakers ignore this evidence, it suggests that they are using administrative burdens to engage indirectly in policymaking about the size and scope of a safety net program.

Shift Burdens from the Individual to the State

Social welfare programs are characterized by conditionality—that is, eligibility criteria that individuals must meet to access benefits. The question is who absorbs the costs of ensuring those criteria are met. The administrative burden literature suggests various approaches to reduce the costs associated with burdens, often by shifting more of the responsibilities of ensuring access to social welfare benefits away from individual beneficiaries and onto the state (Herd et al. 2013).

There are two broad approaches regarding how the state can reduce administrative burden for individuals. First, the state can provide help to people to navigate complicated administrative systems. Caseworkers, for example, provide social welfare beneficiaries significant support navigating the complicated US social welfare system. Low-income individuals trying to access tax benefits designed for them, such as the Earned Income Tax Credit, rely on tax preparers to help navigate that process. The state can pay for or incentivize private actors to offer help, such as health insurance navigators. Of course, private or nonprofit third-party actors can help clients manage this burden, either for a fee or out of a sense of solidarity.

A second approach to reducing burdens is for the state to change the administrative systems that deliver social welfare programs. The simplest approaches are informational nudges, such as notifying people they may be eligible for a program or explaining how to use their benefits. This reduces learning costs for potential clients. This low-cost approach is popular, partly because it is easy to adopt, and partly because it fits with a behavioral-economics perspective on change. However, "the evidence on the effectiveness of low-cost nudges to encourage participation is mixed, with some interventions showing modest effects on take-up but others showing no significant effect or any effect" (National Academies of Sciences, Engineering, and Medicine 2023, p. 135; DellaVigna and Linos 2022).

More fundamental changes to administrative systems are typically required to substantially reduce administrative burdens for the public. Such changes require a will and capacity to change how the government operates, rather than simpler interventions like sending texts to clients. One of the most effective strategies is automatic enrollment—or process changes that move in that direction (Fox, Stazyk, and Feng 2020; Herd et al. 2013). Indeed, our largest and most effective poverty reduction policy, the Social Security retirement program, does this. Individuals are enrolled into the Social Security retirement benefits by default (Herd and Moynihan 2018). The program is very complex in terms of how eligibility and benefits are determined, but feels simple to the user because the state has organized a machinery to record their earnings and payroll tax contributions over the course of their lives. Automatic enrollment (and recertification) effectively reduces, and even eliminates, learning, compliance, and psychological costs. There has also been movement in low-income programs like Medicaid, where people have to annually demonstrate they are still eligible, to automate that "recertification" process when states have the data they need (such as within their tax administrative systems) to see whether individuals are still eligible (Volkov et al. 2024). In short, rather than requiring Medicaid beneficiaries to provide the paperwork and documentation to prove they are still eligible, states internally verify eligibility with their existing administrative data.

Technology: Reducing or Causing Administrative Burden?

Shifting burdens onto the state via automatic enrollment requires a monitoring regime to ensure that any such system actually works to reduce burdens. It is helpful to distinguish between automatic enrollment and renewals as a specific tactic and the more general trend toward automation of administrative processes, which has a mixed record. Indiana tried to automate eligibility processes that had been handled by individual caseworkers (Wu and Meyer 2023). Indiana Governor Mitch Daniels contracted with IBM to use technology to resolve what he characterized as issues of fraud, error, and poor service. The use of automation allowed remote applications and approvals, which reduced compliance costs. But the resulting systems were inflexible and error-ridden, increasing burdens on clients and leading to large decline in enrollments for the Supplemental Nutrition Assistance Program, Temporary Assistance for Needy Families, and Medicaid. The declines were especially large in higher poverty counties, and even when the state abandoned the automated system, the enrollment declines became permanent.

In a critique of the Indiana system, Eubanks (2017) notes that the automated system overwhelmed caseworkers who were unprepared to work with it, removed their knowledge of clients, and denied them discretion to resolve problems. Automation can generate "digital cages," trapping people in virtual processes where they cannot find human support (Peeters and Widlak 2018). Such failures are more likely to occur when the digital interfaces are poorly designed, when outcomes are not monitored by humans to detect errors or manage appeals, and if the political principals and/or designers of the system are indifferent to imposing burdens on the public. Clients and caseworkers experience the costs and errors of inflexible administrative systems, which may take years to detect and fix.

An appropriate degree of skepticism about the limits of technology, and the ways in which it can go wrong, should not diminish attention to the ways that technology can help to reduce burdens in the safety net. Well-designed digital interfaces and back-end processes can reduce learning, compliance, and psychological costs. There are promising tactics to reduce burdens in the context of improvements to digital interfaces (for example, Moynihan et al. 2022; Giannella et al. 2024). The type of automatic enrollment relies on optimized back-end technologies that allow for accurate and efficient data matching for programs like Medicaid (Herd and Moynihan 2023a). In sum, technology is a tool—it can reduce or exacerbate administrative burdens. Increasingly, government has a digital face, meaning the role of technology becomes ever more central to how burdens are experienced.

Street-Level Bureaucrats Can Buffer or Impose Burdens

While technology and automation will play a larger role in the administration of the safety net, understanding the role of caseworkers and other public officials embedded in these systems remains important. To the degree that "street level bureaucrats" (Jilke and Tummers 2018) have discretion, their judgment, behaviors and decisions can alter the experiences and outcomes of clients (Barnes and Henly 2018). For example, individual caseworkers can have significant effects on the probability of approval for the Supplemental Nutrition Assistance Program for the clients they are assigned, even under conditions where the caseworkers appear to have limited discretion to make decisions (Cook and East 2023).

Understanding what motivates these officials and how they use their discretion is relevant for understanding administrative burdens. Audit studies, which reveal that street-level bureaucrats may engage in discriminatory practices when dealing with the public, can usefully consider the role of burdens as a mechanism of discrimination. For example, administrative actors can use their discretion to provide or withhold information to clients (affecting learning costs), to add more demands and offer little help with documentation processes (affecting compliance costs), or offer a more or less welcoming tone (affecting psychological costs) (Olsen, Kyhse-Andersen, and Moynihan 2022).

Program Shifts since Welfare Reform: Lower Burdens, Increased Inequality

An administrative burden approach also offers a mixed assessment of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996—better known as the "welfare reform act"—and its aftermath (Herd and Moynihan 2023a). A major shift in the 1996 act, along with toughening work requirements, was that states received the federal share of Temporary Assistance to Needy Families funds as a block grant—and have considerable freedom to regarding how to spend the money. A growing body of evidence suggests that the high level of discretion that states have, alongside with a wide array of complex programs, has led to substantial burdens in the programs funded by TANF (Herd and Moynihan 2023b). For example, the programs complex array of requirements means that high fractions of participants, as many as 50 percent in Wisconsin, were sanctioned for failing to meet them, with black individuals more likely to be sanctioned (Wu et al. 2006).

At the same time, other aspects of the safety net have grown more quickly, and often feature fewer administrative burdens. For example, before the 1996 welfare reform act, eligibility for the Supplemental Nutrition Assistance Program ("food stamps") had been attached to cash welfare. After welfare reform, claimants had to go through a separate process to claim food stamps (Herd and Moynihan 2018). At first, the new administrative burdens led to substantially lower take-up. But a sustained effort to reduce these burdens was successful (Ganong and Liebman 2018).

The Earned Income Tax Credit and the Child Tax Credit have no separate administrative process that clients must negotiate to access benefits. A good example of their ability to minimize administrative burden occurred during the pandemic, when a temporarily expanded Child Tax Credit proved startlingly successful in reducing poverty and food insecurity (Parolin et al. 2023). Much of this success came from being able to deliver the benefit quickly via the tax system. Moreover, the federal government did not wait for people to apply; it automatically enrolled them. But those with the lowest incomes, who are not legally required to submit taxes, did not benefit from such automatic enrollment and had to apply for the Child Tax Credit separately. Not surprisingly, they benefited from the expansion of that credit at much lower rates than those connected to the tax system. In this way, the full poverty-reducing effects of the Child Tax Credit expansion was blunted by not being able to reach the poorest beneficiaries (Herd and Moynihan 2023a).

Beyond Take-Up and Recertification

The emerging literature on administrative burdens has centered to a great degree on how burdens affect "take-up," or whether people enroll in programs for which they are eligible, and on recertification, which refers to whether people are able to maintain that eligibility when reassessed on a biannual or annual basis. This focus is understandable, but the scope of administrative burdens, as well as their impacts, is broader.

For example, even those who (eventually) access safety benefits may still experience administrative burdens, including loss of time, stress, and anxiety. Reducing the "time tax" that people lose in citizen-state interactions, or negative psychological experiences, has value in and of itself (Lowery 2021).

People can also experience administrative burden when trying to use their safety net benefits, which Barnes (2021) refers to as "redemption costs." For example, the Supplemental Nutrition Assistance Program and Special Supplemental Nutrition Program for Women, Infants, and Children (commonly abbreviated as WIC) have a lengthy list of requirements regarding what food people can and cannot buy, down to the size of the container (Barnes 2021). Medicaid recipients sometimes face obstacles finding and getting an appointment with a doctor that accepts Medicaid (Office of Inspector General 2014). Individuals using Medicaid home and community-based long-term care services face significant hurdles accessing services for which they have been deemed eligible, in large part due to worker shortages linked to low pay (Herd and Johnson 2024).

Those receiving safety net benefits can also experience administrative burdens when targeted by audit mechanisms to verify they have not engaged in fraud (Peeters and Widlak 2023). For example, the Earned Income Tax Credit involves a relatively low-burden application process, but recipients are much more likely to face audit processes that can be difficult to overcome (Herd and Moynihan 2018). Thus, research on administrative burdens considers a broader array of experiences, beyond take-up, better measuring experiences of burden as a topic in its own right across a wider array of venues beyond application and renewal processes (for example, Barnes 2021; Jilke et al. 2024).

It is also the case that administrative burdens likely have downstream effects, beyond immediate participation and use of benefit or service. For example, administrative burdens that hinder access to social safety net programs likely affect long-term health and financial well-being (Herd and Moynihan 2020). Burdens might also shape people's political beliefs, identities, and participation. More negative citizen-state encounters might sap trust in public services, and government more generally (Michener 2018).

In addition to considering the political consequences of administrative burdens, we might consider their political origins. To what degree do they reflect "policymaking by other means"—like a mandate to hold down costs or to benefit some groups over others—rather than a function of bureaucratic inattention (Herd and Moynihan 2018; Herd and Johnson 2024)? Why are persistent administrative burdens tolerated by policymakers, caseworkers, and the mass public (Baekgaard, Moynihan, and Thomsen 2021; Baekgaard, Halling, and Moynihan 2024)?

An International Dimension

While much of the administrative burden research has focused on the safety net in high-income countries, especially the United States, these challenges are likely to be even greater in developing nations, due to a mixture of more limited government capacity to manage burdens and limits on individual access to the resources to overcome burdens (Heinrich 2016; Banerjee et al. 2021). Some of the most insightful empirical work on improving safety net systems comes from such settings, where a variety of experimental interventions have sought to improve the flow of benefits to clients. In some cases, these interventions seek to minimize the risk of corruption among public officials, with the effect of making the administrative processes easier for clients by removing points where officials may siphon off resources or solicit bribes (Nieto-Morales, Peeters, and Lotta 2024). Other cases involve the use of new technologies to bypass outdated or under-resourced administrative processes.

For example, Muralidharan, Niehaus, and Sukhtankar (2016) point to the use of biometric data in anti-poverty programs in India as a means to increase state capacity to reduce "leakage" of benefits, but the program also reduces burdens on clients to provide documentation. Provision of program information to beneficiaries increased subsidies provided to them, reducing learning costs about benefits in a way that allowed citizens to more successfully claim benefits (Banerjee et al. 2018). Help in registering for public health insurance increased take-up in Indonesia, but even with help, people continued to struggle to manage unwieldy processes, unreliable administrative data, and distant in-person offices: "In short, administrative capacity constraints generate hassle costs that suppress demand" (Banerjee et al. 2021, p. 3038). It is not just about capacity. Qualitative work in settings as varied as Pakistan and Mexico points to how citizen understanding of administrative burdens, and strategies to manage them, reflect local cultural norms (Masood and Nisar 2021; Peeters et al. 2018). Such research enriches our understanding of the relationship between state capacity, local culture, and safety net systems, and may build additional conceptual linkages to other studies by clarifying the administrative burden aspect of such work.

Conclusion

If more economists are, as Duflo (2017) urges, to become as useful as plumbers, tinkering and adjusting with existing policy systems, the administrative burden approach offers some useful tools to understand why these burdens matter, how to identify them, and how to fix them. The approach also provides a distinctive orientation for policymakers towards identifying and reducing burdens, viewing them as damaging to program goals, client outcomes, as well as devaluing the time of citizens. Indeed, the language of "administrative burden," as drawn from scholarship and used by policymakers, offers a bridge between these communities.

The administrative burden approach invites a relatively broad research agenda. Some of these questions will be relatively straightforward extensions of behavioral perspectives. For example, why do people not access services they want and desire? What are the heterogeneous effects of current structures? How does this relate to inequality despite seemingly neutral systems? What are ways to make it easier for people to access those services? In particular, what role can technology play as a means to reduce burdens (or in some cases, increase burdens)? Extending take-up analyses to consider the downstream effects of burdens on, for example, public health, as well as a deeper understanding of the political economy of burdens offers an opportunity to extend the research framework.

References

Arbogast, Iris, Anna Chorniy, and Janet Currie. 2022. "Administrative Burdens and Child Medicaid Enrollments." NBER Working Paper 30580.

Baekgaard, Martin, Aske Halling, and Donald Moynihan. 2024. "Burden tolerance: Developing a validated measurement instrument across seven countries." *Public Administration Review*. https://doi.org/10.1111/puar.13835.

Baekgaard, Martin, Pamela Herd, and Donald P. Moynihan. 2023. "Of 'Welfare Queens' and 'Poor Carinas': Social Constructions, Deservingness Messaging and the Mental Health of Welfare Clients." *British Journal of Political Science* 53 (2): 594–612.

Baekgaard, Martin, Donald P. Moynihan, and Mette Kjærgaard Thomsen. 2021. "Why Do Policymakers Support Administrative Burdens? The Roles of Deservingness, Political Ideology, and Personal Experience." *Journal of Public Administration Research and Theory* 31 (1): 184–200.

Banerjee, Abhijit, Amy Finkelstein, Rema Hanna, Benjamin A. Olken, Arianna Ornaghi, and Sudarno Sumarto. 2021. "The Challenges of Universal Health Insurance in Developing Countries: Experimental Evidence from Indonesia's National Health Insurance." *American Economic Review* 111 (9): 3035–63.

Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto. 2018. "Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia." *Journal of Political Economy* 126 (2): 451–91.

Barnes, Carolyn Y. 2021. "'It Takes a While to Get Used to': The Costs of Redeeming Public Benefits." *Journal of Public Administration Research and Theory* 31 (2): 295–310.

Barnes, Carolyn Y., and Julia R. Henly. 2018. "'They Are Underpaid and Understaffed': How Clients Interpret Encounters with Street-Level Bureaucrats." *Journal of Public Administration Research and Theory* 28 (2): 165–81.

Barofsky, Jeremy, Ariadna Vargas, Dinardo Rodriguez, and Anthony Barrows. 2020. "Spreading Fear: The Announcement of the Public Charge Rule Reduced Enrollment in Child Safety-Net Programs: Study Examines Whether the Announced Change to the Federal Public Charge Rule Affected the Share of Children Enrolled in Medicaid, SNAP, and WIC." *Health Affairs* 39 (10): 1752–61.

Bell, Elizabeth, Julian Christensen, Pamela Herd, and Donald Moynihan. 2023. "Health in Citizen-State Interactions: How Physical and Mental Health Problems Shape Experiences of Administrative Burden and Reduce Take-Up." *Public Administration Review* 83 (2): 385–400.

Biden, Joe. 2021a. *Advancing Racial Equity and Support for Underserved Communities through the Federal Government*. Executive Order 13985. Executive Office of the President.

Biden, Joe. 2021b. *Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government*. Executive Order 14058. Executive Office of the President.

Center on Budget and Policy Priorities. 2022. "Policy Basics: Temporary Aid to Needy Families." March 1. https://www.cbpp.org/research/family-income-support/policy-basics-an-introduction-to-tanf.

Christensen, Julian, Lene Aarøe, Martin Baekgaard, Pamela Herd, and Donald P. Moynihan. 2020. "Human Capital and Administrative Burden: The Role of Cognitive Resources in Citizen-State Interactions." *Public Administration Review* 80 (1): 127–36.

Congressional Budget Office (CBO). 2024. "Health Insurance and Its Federal Subsidies: CBO and JCT's June 2024 Baseline Projections." June. https://www.cbo.gov/system/files/2024-06/51298-2024-06-healthinsurance.pdf.

Cook, Jason B., and Chloe N. East. 2023. "The Effect of Means-Tested Transfers on Work: Evidence from Quasi-randomly Assigned SNAP Caseworkers." NBER Working Paper 31307.

Cook, Jason B., and Chloe N. East. 2024. "The Disenrollment and Labor Supply Effects of SNAP Work Requirements." NBER Working Paper 32441.

Cuffey, Joel, Timothy K. M. Beatty, and Elton Mykerezi. 2022. "Work Effort and Work Requirements for Food Assistance among US Adults." *American Journal of Agricultural Economics* 104 (1): 294–317.

Currie, Janet. 2006. "The Take-Up of Social Benefits." In *Public Policy and the Income Distribution*, edited by Alan J. Auerbach, David Card, and John M. Quigley, 80–148. Russell Sage Foundation.

Dague, Laura, and Benjamin Ukert. 2024. "Pandemic-Era Changes to Medicaid Enrollment and Funding: Implications for Future Policy and Research." *Journal of Policy Analysis and Management* 43 (4): 1229–59.

DellaVigna, Stefano, and Elizabeth Linos. 2022. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." *Econometrica* 90 (1): 81–116.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy* 11 (4): 213–48.

Dreyfus, Hannah. 2021. "States Are Hoarding $5.2 Billion in Welfare Funds Even as the Need for Aid Grows." *ProPublica*, December 29. https://www.propublica.org/article/states-are-hoarding-52-billion-in-welfare-funds-even-as-the-need-for-aid-grows.

Duflo, Esther. 2017. "The Economist as Plumber." *American Economic Review* 107 (5): 1–26.

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Fox, Ashley M., Edmund C. Stazyk, and Wenhui Feng. 2020. "Administrative Easing: Rule Reduction and Medicaid Enrollment." *Public Administration Review* 80 (1): 104–17.

Goldstein, Adam, Charlie Eaton, Amber Villalobos, Parijat Chakrabarti, Jeremy Cohen, and Katie Donnelly. 2023. "Administrative Burden in Federal Student Loan Repayment, and Socially Stratified Access to Income-Driven Repayment Plans." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 9 (4): 86–111.

Ganong, Peter, and Jeffrey B. Liebman. 2018. "The Decline, Rebound, and Further Rise in SNAP Enrollment: Disentangling Business Cycle Fluctuations and Policy Changes." *American Economic Journal: Economic Policy* 10 (4): 153–76.

Giannella, Eric, Tatiana Homonoff, Gwen Rino, and Jason Somerville. 2024. "Administrative Burden and Procedural Denials: Experimental Evidence from SNAP." *American Economic Journal: Economic Policy* 16 (4): 316–40.

Gray, Colin. 2019. "Leaving Benefits on the Table: Evidence from SNAP." *Journal of Public Economics* 179: 104054.

Gray, Colin, Adam Leive, Elena Prager, Kelsey Pukelis, and Mary Zaki. 2023. "Employed in a SNAP? The Impact of Work Requirements on Program Participation and Labor Supply." *American Economic Journal: Economic Policy* 15 (1): 306–41.

Guth, Madeline, Patrick Drake, Robin Rudowitz, and Maiss Mohamed. 2023. "Understanding the Intersection of Medicaid Work: A Look at What the Data Say." Kaiser Family Foundation, April 24. https://www.kff.org/medicaid/issue-brief/understanding-the-intersection-of-medicaid-work-a-look-at-what-the-data-say/.

Haeder, Simon F., and Donald Moynihan. 2023. "Race and Racial Perceptions Shape Burden Tolerance for Medicaid and Supplemental Nutritional Assistance Program." *Health Affairs* 42 (10): 1334–43.

Han, Jeehoon. 2022. "The Impact of Snap Work Requirements on Labor Supply." *Labour Economics* 74: 102089.

Heinrich, Carolyn J. 2016. "The Bite of Administrative Burden: A Theoretical and Empirical Investigation." *Journal of Public Administration Research and Theory* 26 (3): 403–20.

Heinrich, Carolyn J., Sayil Camacho, Sarah Clark Henderson, Mónica Hernández, and Ela Joshi. 2022. "Consequences of Administrative Burden for Social Safety Nets that Support the Healthy Development of Children." *Journal of Policy Analysis and Management* 41 (1): 11–44.

Herd, Pamela, Thomas DeLeire, Hope Harvey, and Donald P. Moynihan. 2013. "Shifting Administrative Burden to the State: The Case of Medicaid Take-Up." *Public Administration Review* 73 (S1): S69–S81.

Herd, Pamela, Hilary Hoynes, Jamila Michener, and Donald Moynihan. 2023. "Introduction:

Administrative Burden as a Mechanism of Inequality in Policy Implementation." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 9 (4): 1–30.

Herd, Pamela, and Rebecca A. Johnson. 2024. "Rationing Rights: Administrative Burden in Medicaid Long term Care Programs." *Journal of Health Politics, Policy, and Law.* https://doi.org/10.1215/03616878-11567708.

Herd, Pamela, and Donald P. Moynihan. 2018. *Administrative Burden: Policymaking by Other Means.* Russell Sage Foundation.

Herd, Pamela, and Donald Moynihan. 2020. "How Administrative Burdens Can Harm Health." *Health Affairs*, October 2. https://www.healthaffairs.org/do/10.1377/hpb20200904.405159/.

Herd, Pamela, and Donald Moynihan. 2023a. "Implementing the Expanded Child Tax Credit: What Worked, What Didn't, and How to Move Forward." *Annals of the American Academy of Political and Social Science* 710 (1): 58–74.

Herd, Pamela, and Donald Moynihan. 2023b. "Fewer Burdens but Greater Inequality? Reevaluating the Safety Net through the Lens of Administrative Burden." *Annals of the American Academy of Political and Social Science* 706 (1): 94–107.

Herd, Pamela, and Donald Moynihan. 2024. "Gendered Administrative Burden: Regulating Gendered Bodies, Labor, and Identity." *Journal of Public Administration Research and Theory.* https://doi.org/10.1093/jopart/muae021.

Herd, Pamela, Donald Moynihan, and Amy Widman. 2023. *Identifying and Reducing Burdens in Administrative Processes. Report to the Administrative Conferences of the United States.* Administrative Conference of the United States.

Homonoff, Tatiana, and Jason Somerville. 2021. "Program Recertification Costs: Evidence from SNAP." *American Economic Journal: Economic Policy* 13 (4): 271–98.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. "Long-Run Impacts of Childhood Access to the Safety Net." *American Economic Review* 106 (4): 903– 34.

Jilke, Sebastian, Martin Bækgaard, Pamela Herd, and Donald Moynihan. 2024. "Short and Sweet: Measuring Experiences of Administrative Burden." *Journal of Behavioral Public Administration* 7. https://doi.org/10.30636/gg5h9x60.

Jilke, Sebastian, and Lars Tummers. 2018. "Which Clients Are Deserving of Help? A Theoretical Model and Experimental Test." *Journal of Public Administration Research and Theory* 28 (2): 226–38.

Kelman, Brett, and Mike Reicher. 2019. "At Least 220,000 Tennessee Kids Faced Loss of Health Insurance Due to Lacking Paperwork." *Tennessean*, July 14. https://www.tennessean.com/story/news/investigations/2019/07/14/tenncare-coverkids-medicaid-children-application-insurance-denied/1387769001/.

Lin, Jou Chun. 2023. "Designing Effective Welfare Programs: Evidence from SNAP's BBCE Programs." Unpublished.

Lowrey, Annie. 2021. "The Time Tax: Why Is So Much American Bureaucracy Left to Average Citizens?" *Atlantic*, July 27. https://www.theatlantic.com/politics/archive/2021/07/how-government-learned-waste-your-time-tax/619568/.

Madsen, Jonas K., Kim S. Mikkelsen, and Donald P. Moynihan. 2022. "Burdens, Sludge, Ordeals, Red Tape, Oh My!: A User's Guide to the Study of Frictions." *Public Administration* 100 (2): 375–93.

Martin, Lucie, Liam Delaney, and Orla Doyle. 2023. "Everyday Administrative Burdens and Inequality." *Public Administration Review* 84 (4): 660–73.

Masood, Ayesha, and Muhammad Azfar Nisar. 2021. "Administrative Capital and Citizens' Responses to Administrative Burden." *Journal of Public Administration Research and Theory* 31 (1): 56–72.

McGarry, Kathleen, and Robert F. Schoeni. 2015. "Understanding Participation in SSI." University of Michigan Retirement Research Center Working Paper 2015–319.

Michener, Jamila. 2018. *Fragmented Democracy: Medicaid, Federalism, and Unequal Politics.* Cambridge University Press.

Moffitt, Robert. 1983. "An Economic Model of Welfare Stigma." *American Economic Review* 73 (5): 1023–35.

Moynihan, Donald, Eric Giannella, Pamela Herd, and Julie Sutherland. 2022. "Matching to Categories: Learning and Compliance Costs in Administrative Processes." *Journal of Public Administration Research and Theory* 32 (4): 750–64.

Mueller, Holger, and Constantine Yannelis. 2022. "Increasing Enrollment in Income-Driven Student Loan Repayment Plans: Evidence from the Navient Field Experiment." *Journal of Finance* 77 (1): 367–402.

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why Having Too Little Means So Much.* Henry Holt and Company.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. "Building State Capacity: Evidence from Biometric Smartcards in India." *American Economic Review* 106 (10): 2895–2929.

National Academies of Sciences, Engineering, and Medicine. 2023. *Behavioral Economics: Policy Impact and Future Directions.* National Academies Press.

Ndumele, Chima D., Hannah Factor, Matthew Lavallee, Anthony Lollo Jr., and Jacob Wallace. 2024. "Supplemental Nutrition Assistance Program Work Requirements and Safety-Net Program Participation." *JAMA Internal Medicine.* https://doi.org/10.1001/jamainternmed.2024.5932.

Nichols, Albert L., and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *American Economic Review* 72 (2): 372–77.

Nieto-Morales, Fernando, Rik Peeters, and Gabriela Lotta. 2024. "Burdens, Bribes, and Bureaucrats: The Political Economy of Petty Corruption and Administrative Burdens." *Journal of Public Administration Research and Theory* 34 (4): 481–97.

Office of Inspector General. 2014. *Access to Care: Provider Availability in Medicaid Managed Care.* US Department of Health and Human Services.

Olsen, Asmus Leth, Jonas Høgh Kyhse-Andersen, and Donald P. Moynihan. 2022. "The Unequal Distribution of Opportunity: A National Audit Study of Bureaucratic Discrimination in Primary School Access." *American Journal of Political Science* 66 (3): 587–603.

Parolin, Zachary, Elizabeth Ananat, Sophie Collyer, Megan Curran, and Christopher Wimer. 2023. "The Effects of the Monthly and Lump-Sum Child Tax Credit Payments on Food and Housing Hardship." *AEA Papers and Proceedings* 113: 406–12.

Peeters, Rik, Humberto Trujillo Jiménez, Elizabeth O'Connor, Pascual Ogarrio Rojas, Michele González Galindo, and Daniela Morales Tenorio. 2018. "Low-Trust Bureaucracy: Understanding the Mexican Bureaucratic Experience." *Public Administration and Development* 38 (2): 65–74.

Peeters, Rik, and Arjan Widlak. 2018. "The Digital Cage: Administrative Exclusion through Information Architecture—The Case of the Dutch Civil Registry's Master Data Management System." *Government Information Quarterly* 35 (2): 175–83.

Peeters, Rik, and Arjan C. Widlak. 2023. "Administrative Exclusion in the Infrastructure-Level Bureaucracy: The Case of the Dutch Daycare Benefit Scandal." *Public Administration Review* 83(4): 863–77.

Ratcliffe, Caroline, Signe-Mary McKernan, Laura Wheaton, Emma Kalish, Catherine Ruggles, Sara Armstrong, and Christina Oberlin. 2016. *Asset Limits, SNAP Participation, and Financial Stability.* Urban Institute.

Ray, Victor, Pamela Herd, and Donald Moynihan. 2023. "Racialized Burdens: Applying Racialized Organization Theory to the Administrative State." *Journal of Public Administration Research and Theory* 33 (1): 139–52.

Shepard, Mark, and Myles Wagner. 2022. "Do Ordeals Work for Selection Markets? Evidence from Health Insurance Auto-Enrollment." NBER Working Paper 30781.

Sommers, Benjamin D., Lucy Chen, Robert J. Blendon, E. John Orav, and Arnold M. Epstein. 2020. "Medicaid Work Requirements in Arkansas: Two-Year Impacts on Coverage, Employment, and Affordability of Care." *Health Affairs* 39 (9): 1522–30.

Sommers, Benjamin D., Anna L. Goldman, Robert J. Blendon, E. John Orav, and Arnold M. Epstein. 2019. "Medicaid Work Requirements—Results from the First Year in Arkansas." *New England Journal of Medicine* 381 (11): 1073–82.

Sunstein, Cass R. 2021. *Sludge: What Stops Us from Getting Things Done and What to Do about It.* MIT Press.

Sunstein, Cass R. 2022. "Sludge Audits." *Behavioural Public Policy* 6 (4): 654–73.

Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Penguin.

Tolbert, Jennifer, and Bradely Corallo. 2024. "An Examination of Medicaid Renewal Outcomes and Enrollment Changes at the End of the Unwinding." KFF, September 18. https://www.kff.org/medicaid/issue-brief/an-examination-of-medicaid-renewal-outcomes-and-enrollment-changes-at-the-end-of-the-unwinding/.

Trump, Donald. 2018. "Reducing Poverty in American by Promoting Opportunity and Economic Mobility." Executive Order 13828. https://www.federalregister.gov/documents/2018/04/13/2018-07874/reducing-poverty-in-america-by-promoting-opportunity-and-economic-mobility.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.

US Executive Office of the President. 2023. *Tackling the Time Tax: How the Federal Government is Reducing Burdens to Accessing Critical Benefits and Services.* Executive Office of the President.

US Office of Information and Regulatory Affairs. 2022. "Improving Access to Public Benefits Programs through the Paperwork Reduction Act." Office of Management and Budget Working Paper M-22-10.

US Office of Management and Budget. 2021. *Study to Identify Methods to Assess Equity: Report to the President*. Executive Office of the President.

US Office of Management and Budget. 2022. "Section 280—Managing Customer Experience and Improving Service Delivery." OMB Circular A-11.

Volkov, Eden, Amelia Whitman, Nancy De Lew, and Thomas Buchmueller. 2024. "Evaluating Medicaid Strategies to Streamline Ex Parte Renewals." Data Point, ASPE Office of Health Policy, November 14. https://aspe.hhs.gov/sites/default/files/documents/d1c68327569498e86e98764e536b74f4/ex-parte-renewal-data-point.pdf.

Wu, Chi-Fang, Maria Cancian, Daniel R. Meyer, and Geoffrey L. Wallace. 2006. "How Do Welfare Sanctions Work?" *Social Work Research* 30 (1): 33–50.

Wu, Derek, and Bruce D. Meyer. 2023. "Certification and Recertification in Welfare Programs: What Happens When Automation Goes Wrong?" NBER Working Paper 31437.

Zeckhauser, Richard. 2021. "Strategic Sorting: The Role of Ordeals in Health Care." *Economics Philosophy* 37 (1): 64–81.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 151–180

Getting Infrastructure Built: The Law and Economics of Permitting

Zachary Liscow

The US permitting regime for infrastructure projects has undergone a sea change since the 1950s. In one prominent episode, between 1948 and the mid-1960s, Robert Moses, New York's unelected city construction coordinator, bulldozed through the Bronx to build the Cross-Bronx Expressway, with apparent disregard for the thousands of people—many children of refugees—whom the construction displaced. Moses did not consult with them and failed to make even minor adjustments that could have significantly reduced disruption (Caro 1974). Across the United States during this time, projects with large effects on the human environment were undertaken with virtually no review.

The modern permitting regime is quite different. In 2001, the developer Cape Wind began environmental review on an offshore wind energy project in Nantucket Sound (Kimmell and Stalenhoef 2011). After opposition from parties including those (like then–Senator Edward Kennedy) who would see the turbines from their homes, Cape Wind received state permits in 2009 and federal permits in 2010, almost ten years into the project. Multiple lawsuits then continued. Largely because of the long delay, Cape Wind was shuttered after costing the developer $100 million (Seelye 2017).

In the 1960s, the United States did big things with little public consultation. Now, even smaller things can be held up by small opposition groups.

Although concerns about the delays and costs of permitting have been simmering for years, recent developments affecting energy and transportation infrastructure have brought interest in the topic to a full boil. Several recent laws have provided hundreds of billions of dollars of federal funding for a surge of large

■ *Zachary Liscow is Professor, Yale Law School, New Haven, Connecticut. His email address is zachary.liscow@yale.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20221347.

building and infrastructure projects, including the Infrastructure Investment and Jobs Act of 2021, the CHIPS and Science Act of 2022, and the Inflation Reduction Act of 2022. Many of these projects must go through a permitting process.

How can the United States build infrastructure at reasonable speed and cost, without undue harms to communities, the environment, and deliberative processes? Of course, permitting is just one potential cause of high costs and slow infrastructure construction among many, including weak competition and low state capacity in procurement (Liscow, Nober, and Slattery 2023; Liscow 2024), excessive proceduralism (Bagley 2021; Pahlka 2023), poor project management (Flyvbjerg and Gardner 2023; Goldwyn et al. 2023), and legal issues specific to certain types of infrastructure (such as those leading to the long queues for new electricity transmission). But the current US permitting regime contributes to the problem as well.

This article explores the effects of permitting (that is, requirements for government approval) on physical infrastructure construction, focusing on energy and transportation infrastructure. In addition to the importance of infrastructure for economic growth (Allen and Arkolakis 2014) and economic mobility (Chetty and Hendren 2018), the United States needs to build rapidly an immense amount of infrastructure for electricity generation and long-distance transmission to make a transition to carbon-free energy. However, current development timelines are very slow; for example, the Solar Energy Industries Association (n.d.) estimates that bringing a utility-scale solar project from inception to commercial service takes six years, four of which are spent on planning and permitting. Perversely, the current permitting regime may be disproportionately stifling green energy projects compared to fossil fuel projects, as green energy is so much of proposed new development (Energy Technologies Area, Berkeley Lab 2023). Also troubling is that US transportation infrastructure appears to be very high-cost compared with other countries. For example, it costs three times as much on average to build urban transit in the United States as in other rich countries (NYU Transit Costs Project 2022), and the cost of building a mile of US interstate highway tripled in the latter part of the twentieth century as permitting requirements increased (Brooks and Liscow 2023).

I begin with a discussion of how infrastructure permitting works in the United States, with a focus on the National Environmental Policy Act (NEPA). While NEPA is independently important because of its national scope and role in a large share of big infrastructure projects, the discussion is also meant as a case study for the broader issue of permitting infrastructure, much of which is not impacted by NEPA. I then turn to the admittedly limited evidence on costs and benefits of permitting. On the cost side, environmental review has become considerably lengthier in recent decades, and at least some infrastructure costs have greatly increased since the passage of NEPA, though evidence of causality remains elusive. On the benefits side, while case studies suggest that NEPA has curbed some of the worst abuses, more systematic data on benefits are scanty.

For insight into potential underlying causes, I turn to international comparisons of infrastructure permitting and construction, which suggest that the United States experiences higher costs, more parochial planning, and below-average environmental performance. These patterns partly trace to "adversarial legalism" in the United States, in which authority is diffuse and disputes are often resolved by litigation.

Different permitting regimes can be categorized along on two dimensions: the *power* of the executive branch to make decisions, especially versus the judiciary, and the *capacity* of the executive branch to plan. The US permitting regime—federal, state, and local—features low executive branch power, partly because of powerful judicial review, and low planning capacity, partly because of relatively weak internal staffing and the split responsibilities in our federal system. I sketch a menu of possible reform along these dimensions of executive power and planning capacity.

In addition, I make a tentative case that moving along both dimensions together could represent a complementary "green bargain" by improving broadly representative front-end participation, but reducing back-end litigation: a higher-capacity executive branch could be better-equipped to wield more power. Such a set of reforms could result in improvements overall in efficiency (faster construction and lower costs), the environment (lower greenhouse gas emissions), and democracy (improved public participation, especially by disadvantaged groups, and outcomes more reflective of public preferences).

How Permitting Works in Practice

The National Environmental Policy Act

President Richard Nixon signed the National Environmental Policy Act in 1970. (The law is almost always referred to as NEPA, and that is how I will refer to it in this essay.) For major projects located on federal lands or waters, built with federal funding, or subject to federal regulatory approval, the law requires a review of the effects on the "environment," including nature and people. NEPA applies to government and private parties. For a transportation project, the party asking for approval is typically the state department of transportation or a local transit agency. For energy infrastructure, it is often a private developer. These parties do considerable planning, often with some public consultation, before requesting a permit from the federal government.

A project can be subject to three levels of environmental review under NEPA: categorical exclusion, environmental assessment, and environmental impact statement.

Categorical exclusions apply to projects that "normally do not have a significant effect on the human environment" (40 CFR § 1508.1[d]). Agencies maintain a list of activities that generally qualify for categorical exclusion, and such actions do not require detailed environmental analysis. A categorical exclusion typically applies to highway repairs, for example.

An environmental assessment, used when the significance of environmental impacts is unknown but not expected to be large, is intended to be a concise review of the environmental consequences of a project. In practice, environmental assessments are often not concise, such as the atypically long 4,007-page one for New York City's scheme to have a "congestion pricing" zone in part of Manhattan. But if an environmental assessment makes a "finding of no significant impact," the agency can allow the project to move forward. An average of about 12,000 environmental assessments are prepared each year (Council on Environmental Quality 2016).

A full environmental impact statement is used if significant impacts are anticipated from the outset of the process or (less typically) if an environmental assessment determines that environmental impacts are significant. Such a statement involves a detailed evaluation of the proposed project and plausible alternatives, including public involvement and comments from citizens, public interest organizations, industry, and other federal, state, and local government agencies. Over 2010–2018, there were on average 155 final environmental impact statements per year (Council on Environmental Quality 2020). Environmental impact statements are done for new highways and new rail lines, for example.

The requirements imposed by NEPA are principally procedural, rather than substantive. As a matter of process, environmental impacts and alternative designs must be considered. However, NEPA does not impose substantive rules on what the size of those impacts must be or mandate that a certain alternative be adopted.

Agencies at least nominally have discretion to determine the factors they consider and the methods they employ in their assessment. However, the agency can be sued for not producing a thorough enough document. Plaintiffs typically seek to nullify the permit or other agency action, which usually stops the project from proceeding if successful, because there would need to be a new environmental review. Projects are often paused during litigation as well.

The Administrative Procedure Act provides that assessments conducted under NEPA are reviewable by federal courts and shall be set aside if they are "arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law." Under the prevailing interpretation of this broad language, the court's job is not so much to determine whether the agency made the "right" decision, but instead to ensure that the agency did not rely on factors that Congress did not intend it to consider, nor entirely fail to consider an important aspect of the problem, nor arrive at a conclusion that is so implausible that it cannot be ascribed to a mere difference in view or the product of agency expertise. However, courts have applied this ostensibly deferential standard with a "hard look doctrine" that effectively offers judges considerable power to second-guess the agency and to consider the substance as well as the process of the agency's decision (Miles and Sunstein 2008). So, while an agency may have discretion on paper in conducting a NEPA assessment, judicial review can substantially limit that discretion in practice.

Other Permitting Requirements in the United States

NEPA is just one part of permitting in the United States; there are many others. Many of these other permits are substantive, limiting impacts on endangered species or banning building in certain places, for example; the focus of this essay is the procedurally oriented permitting used in environmental reviews.

First, other federal laws can also impose permitting requirements. In some cases, NEPA permitting requirements "piggyback" onto the requirements of other agencies; in other cases, some activities may be exempt from NEPA, as when many actions of the Environmental Protection Agency are covered only by the Clean Water Act.

Furthermore, a wide range of state and local permits may be required, for everything from water districts to zones of special historical importance. Consider an illustrative example of building a 30-turbine wind farm on private land in Wyoming (Wyoming Renewable Energy Coordination Committee 2022). Although the wind farm is not on federal land, it is in the migration path of birds protected by the Endangered Species Act, which requires applying for a federal Incidental Take Permit. The transmission lines cross a federal highway, triggering NEPA. At the state level, the plan needs approval under the Wyoming Industrial Development and Siting Act. It also needs approval from the Board of County Commissioners, and 11 of Wyoming's 23 counties impose regulations beyond the state statutory minima. Although these three levels of authorities are supposed to coordinate, the processes are distinct. Also, construction permits are just the first stage: when connecting the wind farm to the grid, the Federal Energy Regulatory Commission has further requirements.

These state and local permitting regimes share many of the same objectives as NEPA, and some rules seek to reduce duplicative efforts. However, the requirements of the federal and state processes may differ. For example, the federal permitting system could require only the more limited environmental assessment, while the state level requires the equivalent of a fuller environmental impact assessment (or vice versa).

For many types of developments, such as wind and solar developments that are not on federal land or water, do not have endangered species issues, and are not receiving federal funding, *only* state and local permitting matter. However, NEPA is likely to play a role in many of the interstate and interregional transmission projects necessary to make the green energy transition, particularly in the West where crossing federal land is all but guaranteed. NEPA is critical for offshore wind farms, given that the average installation is sited in federal waters beyond the range of state jurisdiction. Also, many new clean energy projects that receive federal funding from the Inflation Reduction Act of 2022 will fall under NEPA requirements, even if not sited on federal lands. Most major surface transportation projects receive federal funding, making them subject to NEPA.

Evidence on Costs and Benefits of the Current Permitting Regime

There are tradeoffs in the intensiveness of the procedures involved in permitting, which can range between the low-intensity process used for 1950s-era highways and the high-intensity processes of today, such as that for offshore wind described earlier, with strict government procedures, lengthy documents, and frequent litigation. The NEPA permitting regime has helped largely stop the mass destruction of areas within US cities—without consulting residents—to build infrastructure, but appears to have now slowed new construction.

On the one hand, a more intensive process can have costs. A more intensive process can delay a project's benefits, especially because delays can build on themselves due to shifting politics, technologies, or rules. Sometimes projects can be cancelled or not considered altogether. Intensive permitting can also lead to higher construction costs, as project planners proactively or reactively build in more expensive ways to respond to permitting requirements or subsequent litigation.

On the other hand, more intensive permitting can have benefits. One benefit can be to help the local environment through scrutiny, public involvement, and mitigation of impacts, as well as from making construction harder. Another benefit in principle is that public participation and empowering potentially impacted parties can produce outcomes reflecting public preferences. This public participation may also directly have "procedural" benefits by helping the public feel involved and helping the ultimate decision seem legitimate.

The optimal rule is challenging to specify because of the diversity of values at stake. But, at a high level, the key tradeoff is between getting needed infrastructure built and having sufficiently broad public input to protect the environment and social values. Quantitative evidence on the costs and benefits of the current permitting regime under NEPA is limited, but some evidence is available.

Costs

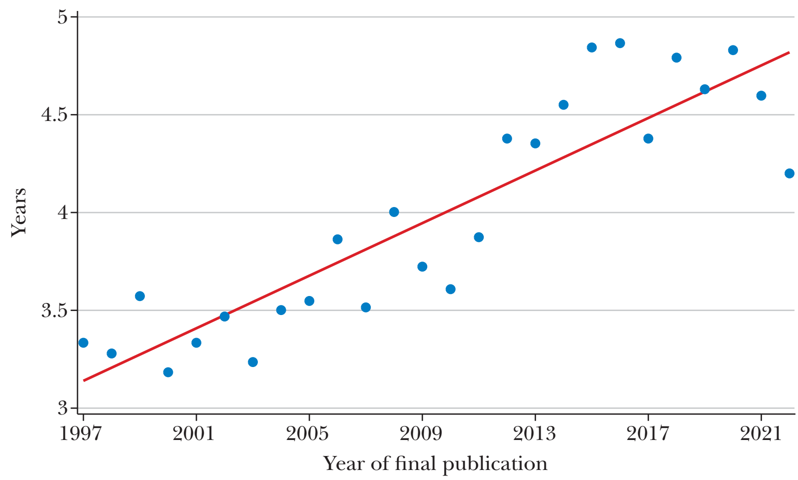
The direct financial costs of environmental review are typically comparatively small. The US Department of Energy reports that its average contractor cost for an environmental impact statement between 2003 and 2012 was $1.4 million, and a 2003 Council on Environmental Quality report estimated that a typical environmental impact statement cost between $250,000 and $2 million (US Government Accountability Office 2014). These costs are for projects that often cost hundreds of millions, if not billions, of dollars.

However, environmental review can lead to much higher costs in other ways. One issue is the incentive to build projects in a more expensive way, either preemptively as "defensive medicine" against potential litigation or in response to litigation. For example, to ease local opposition to a permit, a rail line may be built elevated on pylons to avoid dividing farmland; an urban transit line may be built far underground with a tunneling machine rather than digging up the street to build it less far underground; or a highway may be built with sound walls, depressed underground, or with plazas on top (Brooks and Liscow 2023). Sometimes, projects can become so slow and expensive that they are cancelled—or not considered altogether. Indeed, this is often the goal of project opponents.

Environmental review can also lead projects to proceed more slowly, delaying benefits and increasing costs. This is particularly important because experts say that the time to build energy transmission and generation is significantly slower than needed for the green energy transition (Jenkins et al. 2022).

The duration of environmental review is one indication of delay. One duration measure is the number of days between an agency formally notifying the public of its intent to conduct an environmental review and completing the review. This period includes time for public review and comment. Across all federal agencies, the average final environmental impact statement in 2022 was lengthy, taking 4.2 years

Figure 1
Environmental Impact Statement Preparation Time



Source: DeWitt and DeWitt (2008), extended courtesy of the authors.
Note: Figure 1 shows the average preparation time of all final and final supplemental environmental impact statements (EISs) listed by the Environmental Protection Agency for which a notice of intent is also listed. Preparation time is measured as the number of calendar days between the notice of intent and the final EIS. Year of final publication is when the final EIS is listed in the Federal Register.

to prepare (Nicholson 2022).[1] (Unfortunately, parallel data on the more numerous but less intensive environmental assessments are weak.)

Figure 1 plots the average preparation time for an environmental impact statement across all federal agencies over 1997–2022. Average preparation time generally increased substantially over this period (by about a year-and-a-half), though it decreased a little in 2021 and 2022. Notably, these timelines do not include parts of the project other than permitting, such as initial planning and possibly early public feedback on the front end, as well as litigation and construction on the back end.

Page counts of environmental impact statements are another relevant measure. I collected a random sample of 50 final environmental impact statements from 1977 and 1978, which had an average length of 270 pages without the appendix and 414 pages with it. In comparison, final environmental impact statements between 2013 and 2017 had an average length of 668 pages excluding appendices (a two-and-a-half-fold increase), and an average of 1,703 pages including appendices—a four-fold increase (Council on Environmental Quality 2019).[2] This large increase may be due to a rachet effect in which, once a court expects a given analysis, subsequent environmental impact statements also include that analysis.

[1] The sample includes 95.4 percent of all final environmental impact statements published that year (Nicholson 2022).
[2] Between 2013 and 2017, median length was 445 pages without the appendix, and 962 pages with it.

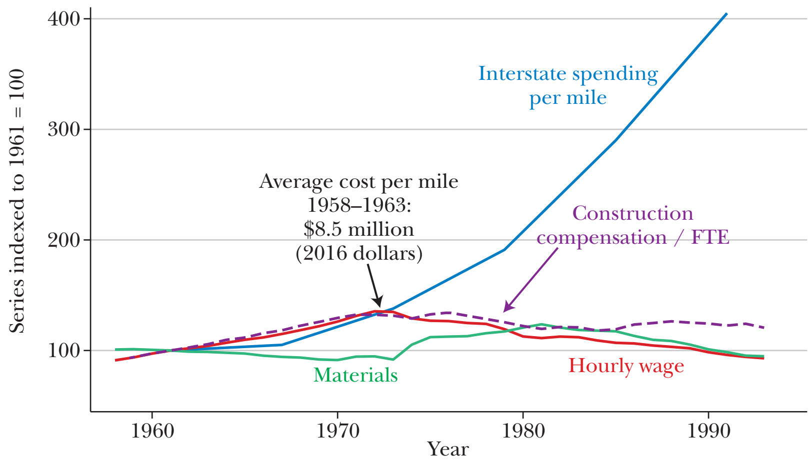
In the experience of E. Donald Elliott, a former General Counsel for the Environmental Protection Agency, most of the detail in environmental review statements is prompted by a desire to address any conceivable issues that might be raised in litigation (Howard 2015). Many government experts agree (US Government Accountability Office 2014; Council on Environmental Quality 1997).

After a final environmental assessment or environmental impact statement is released, the agency may be sued. Many projects have no litigation. In a study of the 355 large infrastructure projects that issued an environmental impact statement between 2010 and 2018, Bennon and Wilson (2023) find that on average 28 percent are litigated, with 89 percent of cases claiming a violation of NEPA. The litigation rates are 64 percent for solar projects, 50 percent for pipelines, 38 percent for wind energy, 33 percent for heavy rail transit, 31 percent for transmission, 29 percent for high-speed rail, and 26 percent for new highways. Of course, the absence of litigation does not imply that the threat of litigation had no impact.

When litigation occurs, it can add substantial time. In a sample of about 300 NEPA cases from 2001 to 2015 (many involving multiple court decisions), Adelman and Glicksman (2018) find a median litigation duration of 23 months, with 75 percent of cases resolved in 39 months and 90 percent in five years. Of the cases filed, 62 percent of nonpending cases filed under NEPA between 2001 and 2011 resulted in judgment for the agency defendant or were dismissed without settlement (author analysis of Council on Environmental Quality n.d., taking the average of yearly rates). When the plaintiff prevailed "on at least one claim, 50% of the cases are resolved within 2.5 years; 75% resolved within about 4.3 years; and 90% of the cases are resolved within 6.2 years." When the government prevailed, the median duration was 18 months, with 75 percent of cases resolved in 36 months. So, even when the project opponent loses in court, the delay can be considerable.

As delays involved in permitting have risen over time, infrastructure costs have risen as well. The costs of interstate highways are suggestive of trends for other forms of infrastructure. Real spending per mile on interstate highways more than tripled between the 1960s and the 1980s—and building in more expensive places does not account for the increase (Brooks and Liscow 2023). Figure 2 shows these costs, indexed to 100 in 1961. The figure also shows that changes in construction wages and materials prices do not explain the increase. Several pieces of evidence link the cost increase to the rise of "citizen voice," in which the public can have a more direct impact on government decision-making. Rising land use litigation statistically explains about one-quarter of the increase in costs. There is an inflection point around 1970, when NEPA passed. Before 1970, there is no relationship between income levels and cost of highways across locations, but a strong positive relationship arises afterward, suggesting that well-heeled communities took more advantage of the changes—perhaps owing to greater resources—to demand more expensive highways. And, after 1970, highways became "wigglier" and more likely to have expensive structures (like noise walls), possibly to placate recalcitrant residents.

Figure 2
Average Spending/Mile to Build New US Interstates, Construction Wages, and Materials Prices over Time



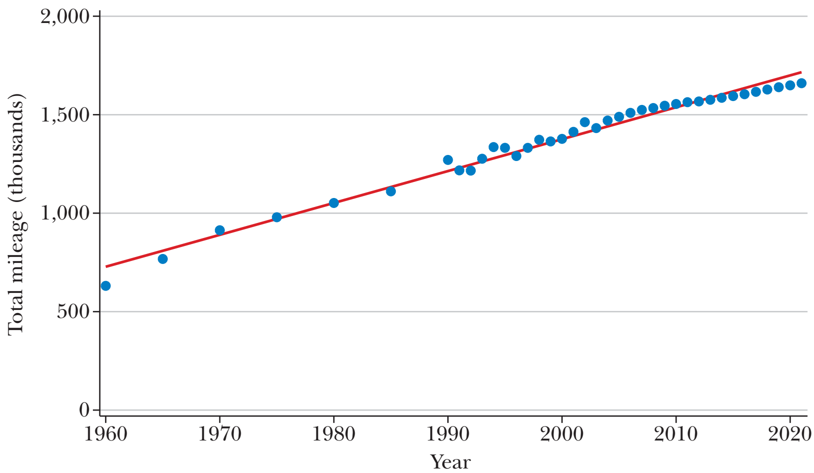
Source: Brooks and Liscow (2023).
Note: This figure shows interstate spending per mile (measured in six-year periods), construction compensation per full-time equivalent employee, construction materials prices, and construction hourly wage in the United States over time. All series are in 2016 dollars, then indexed to 1961.

Benefits

There are large literatures on the substantial benefits of environmental regulation, some of which operate via permitting; for example, a symposium in the Fall 2019 issue of this journal on the 50th anniversary of the Clean Air and Water Acts provides an overview (Currie and Walker 2019; Schmalensee and Stavins 2019; Keiser and Shapiro 2019). These statutes mostly work through imposing rules on behavior, like emissions levels or technology adoption. But there is little well-identified work specifically on NEPA permitting, which works mostly as a *procedural* constraint (Emerson and Baldwin 2019). Ample case studies do suggest improved outcomes from the statute (Pepper 2015). But critics argue that it is difficult to disentangle the benefits specific to NEPA from those stemming from other environmental laws (Potter, Datta, and Stapp 2022).

Those who view NEPA as a way to slow construction of fossil fuel infrastructure can point to some high-profile cases, like the proposals for the Keystone XL pipeline across Montana, South Dakota, and Nebraska (NRDC 2021) or the Atlantic Coast Pipeline that would have run from West Virginia to Virginia and North Carolina (Hinkle and Richardson 2022), in which lawsuits played a role in blocking construction until the projects died. However, as Figure 3 shows, the total mileage of natural gas pipelines has continued to grow at roughly the same rate over time (Bureau of Transportation Statistics 2023). This figure shows total mileage, which includes shutdowns of older segments, and so the amount of new construction is even more

Figure 3
Total Gas Pipeline Mileage in the United States, 1960–2021



Source: Bureau of Transportation Statistics (2023).
Note: Total mileage includes distribution mains, transmission pipelines, and gathering lines. It excludes service pipelines. Data are not adjusted to common diameter equivalent. The line of best fit is: Mileage = 16 * Year – 31,000.

than is shown. Of course, pipeline mileage might have grown even more quickly in the absence of the permitting rules, especially given the shale boom in recent years.

In terms of "democratic" benefits of permitting, there is limited systematic evidence on who takes advantage of permitting laws by submitting comments or suing under NEPA, although there is some evidence available in similar arenas. For example, between 2001 and 2013, public interest groups made up at least the plurality, and often the majority, of plaintiffs in NEPA cases (Council on Environmental Quality n.d.). But "public interest groups" are defined broadly and can include, for example, local "not in my backyard" groups. In the context of regulatory rulemaking, Wagner et al. (2021) find that those participating are mostly companies and trade associations. In the housing development context, Einstein, Palmer, and Glick (2019) find that citizen participants in planning board and zoning board meetings speaking on the construction of housing units are more likely to be older, male, longtime residents, voters in local elections, and homeowners; these participating individuals overwhelmingly oppose the construction of new housing, more so than the general public. It is possible that those making comments and those litigating in the context of infrastructure permitting could skew similarly. After all, the courts and bureaucratic processes are often costly to access successfully, giving a leg up to well-resourced but nonrepresentative groups. For example, in the case of Cape Wind mentioned in the introduction, much of the opposition stemmed from the Alliance to Protect the Nantucket Sound, "a group funded by ultra-wealthy local residents" (Dourado 2023).

However, the permitting regime did seem to empower central urban neighborhoods that protested against interstate highway construction in the 1950s and 1960s. Brinkman and Lin (2024) estimate high welfare costs of highways, especially in central-city neighborhoods. They also find that neighborhood population losses were greatest where central-city freeways hewed most closely to route plans based on historical routes—and were therefore less likely to account for modern neighborhood conditions. As permitting arrived, this rigid, costly adherence to planned freeway routes fell, and the decline was "especially sharp in central neighborhoods" and consistent with "the timeline of policy changes that ceded more power to neighborhood groups" (p. 1270).

Looking ahead, it is important to keep in mind what infrastructure will need permits in the coming decades. Among projects seeking to connect to the electricity grid, 95 percent of the capacity is solar, battery storage, or wind (Energy Technologies Area, Berkeley Lab 2023). This is a dramatic change from the current electricity supply, which is 60 percent fossil fuel and 40 percent zero-carbon-emission (18 percent nuclear and 22 percent renewables, much of that hydroelectric), and an even bigger change from when Congress passed NEPA in 1969, when 81 percent of the electricity supply was fossil fuel and only 19 percent was zero-emission (1 percent nuclear and 18 percent renewables, mostly hydroelectric) (US Energy Information Administration 2023).

US Infrastructure and Permitting in International Context

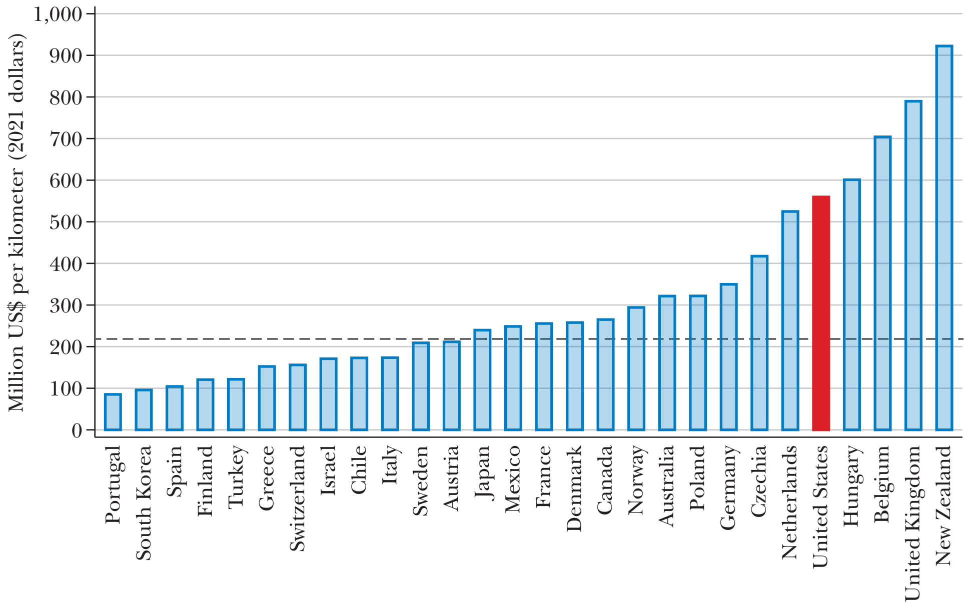
Looking at cross-country differences relevant to permitting—including construction costs, environmental performance, and civic participation—suggests problems that can arise in the distinctive US system of "adversarial legalism," with high levels of litigation and diffuse authority. US permitting is one piece that reflects this broader tendency.

Comparing Outcomes across Countries

Building urban transit megaprojects is far more expensive in the United States, compared most other countries. As Figure 4 shows, spending is about $560 million (in 2021 dollars, adjusted for purchase price parity) per kilometer in the United States, the fifth-highest rate in a sample of 28 OECD countries (author calculations from NYU Transit Costs Project 2020). When adjusting for how little tunneling is done in US projects, the United States may have the highest costs in this group. The horizontal dashed line shows an OECD average of about $200 million per kilometer, a little over one-third the US cost. This OECD average is very close to the global average, computed using all 59 countries for which I have available data. Building in the United States is almost six times as expensive as in Portugal, South Korea, and Spain. It is also worth noting that the five Anglo countries are all in the top dozen most expensive countries, suggesting some importance for legal origins.

Similarly, we can compare the cost of building new highways over time and across countries, combining hand-collected and World Bank data used by Brooks

Figure 4
Average Spending/Kilometer on Urban Transit across Countries



Source: Author calculations, using data from NYU Transit Costs Project (2020).
Note: This figure plots the mean cost per kilometer of urban rail projects. Costs include all construction and construction-related expenditures, but do not include rolling stock or sales taxes. Mean cost per kilometer for each country is adjusted for purchase price parity, then converted to 2021 dollars. This mean is weighted by project length. The dashed, horizontal line displays the overall kilometer-weighted mean across all countries. There are 237 projects total. Sample includes all OECD countries with data.

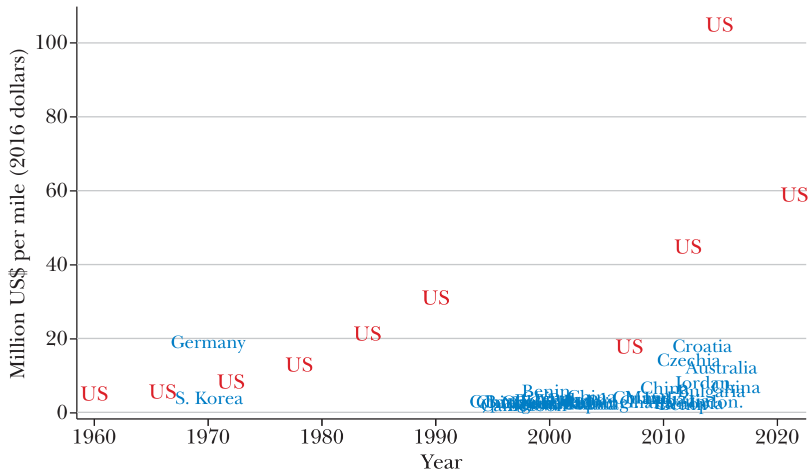
and Liscow (2023), plus more recent data from the Federal Highway Administration's "Major Projects" database. Though the data are sparse, US interstate highways built in the 1980s and 1990s were more expensive (in real terms) than those built elsewhere—and US highways built since 2010 are far more expensive than highway projects elsewhere in the world at any time, as illustrated in Figure 5.

Of course, costs are not the only outcome that matters. It is worth seeing what international data can say about other dimensions that matter, acknowledging that far more is at play than how permits are issued.

The United States is below the OECD average in terms of environmental quality. The Yale Center for Environmental Law & Policy's Environmental Performance Index is constructed to compare countries on climate change performance, environmental health, and ecosystem vitality. The OECD average is 58, while the United States is at 51, at the twenty-fifth percentile between Poland and Hungary (Wolf et al. 2022).

Finally, country experts surveyed by the Varieties of Democracy (V-Dem) dataset rank the United States substantially above average—at the seventy-fifth percentile

Figure 5
Spending per Mile on Highways across Time and Countries



Source: US data after 2000 are from Federal Highway Administration (2024); these "Major Projects" are over $500 million in total cost that have federal assistance. Otherwise data are from Brooks and Liscow (2023), Figure A13.
Note: Data are binned into six-year periods; include only completed new, non-gravel highway or freeway construction; and are mileage-weighted. Project costs from other countries are converted to USD using a nominal exchange rate (as there is not a purchase price parity adjustment series that captures all these years and countries); then this value is put into 2016 dollars using CPI-U.

among OECD countries—in its degree of participation in "political processes, electoral and non-electoral" (Coppedge et al. 2023, p. 52).

Explaining the Cross-Country Differences

I describe here four salient differences between the United States and other countries that can help to explain cross-country differences in permitting: (1) federalism; (2) litigiousness; (3) weak public sector planning capacity; and (4) the presidential system. These ingredients together add up to the US system of "adversarial legalism," in which policy tends to be made through litigation and in a decentralized, fragmented way, rather than through a strong executive—and this broader approach is reflected in permitting.

1) Federalism. Federalism in the United States involves an unusually high degree of decentralization and overlapping authority across different levels of government, which can give project opponents more opportunities to object, increase the layers of procedure to go through, and reduce coordination across space—but may also reflect local preferences better. The US Constitution does not give the federal government an explicit role in infrastructure development—only in "interstate commerce" or "national defense." The level of decentralization is then further expanded by delegations to local governments (Briffault 1990). In contrast, the

Table 1
Cross-Country Measures of Litigation

	Australia	Canada	France	Japan	England	US
Suits filed (per 100,000 people)	1,542	1,450	2,416	1,768	3,681	5,806
Judges (per 100,000 people)	4.00	3.3	12.47	2.83	2.22	10.81

Source: Ramseyer and Rasmusen (2010).

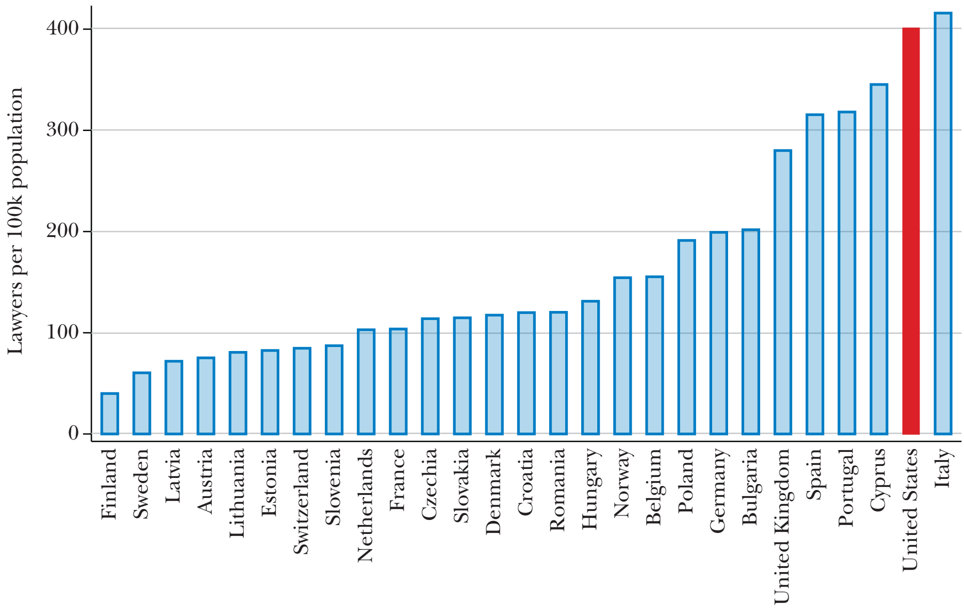
1958 French constitution makes the central government responsible for developing national railways and the National Highway System (European Committee of the Regions n.d.). France, like many European countries, publishes overall transport schema outlining maintenance and new investments (Marshall 2011).

Decentralization in planning is reflected in employment statistics. Of the country's 39,880 urban and regional planners, 28,260 work for local governments, 4,200 work for state governments, and only 870 work for the federal government (US Bureau of Labor Statistics 2023). That is 0.53 percent, 0.23 percent, and 0.04 percent of each workforce, respectively. Federalism fragments US permitting, because more authorities need to be satisfied when crossing multiple jurisdictions (Gordon and Schleicher 2015).

The US permitting system seems especially fragmented relative to those of the European Union, Japan, and New Zealand (Schumacher 2019). For example, gas and electricity companies in Europe face an EU requirement to cooperate with the national government to produce transport/transmission plans (Marshall 2011). While Canada (like the United States) leaves regulation of energy production to provinces, decisions for interprovincial energy transport are left to the federal government (Coleman 2019). Provinces can have input through "cooperative federalism," but cannot veto or "frustrate" the project (Coleman 2019). In contrast, in the United States, large-scale renewable energy installations face at least 121 local policies across 31 states that now restrict renewables development (Eisenson et al. 2024).

2) Litigiousness. Alexis de Tocqueville (1835) wrote, "There is hardly any political question in the United States that sooner or later does not turn into a judicial question." As shown in Table 1, the United States has significantly more suits filed per capita than the other five countries in the sample (Australia, Canada, France, Japan, and England), and far more judges than the other countries except for France (Ramseyer and Rasmusen 2010). Among European countries of over a million people on which I have data, all except for Italy have fewer lawyers per capita than the United States, as shown in Figure 6. The United States has about 400 lawyers per 100,000 people, much higher than common in most western European countries: about 50 percent higher than the United Kingdom, double the ratio in Germany, quadruple the ratio in France. The Nordic and Baltic countries commonly have between 40 and 83 lawyers per 100,000 people.

Figure 6
Lawyers per Capita



Source: Number of lawyers in the United States comes from the American Bar Association (2022)'s measure of resident active attorneys. Number of lawyers for European countries comes from the Council of Bars and Law Societies of Europe (2021), which collected totals from the bars and law societies of individual member countries. Population estimates for 2020 come from World Bank (2023).
Note: This figure displays the number of lawyers per 100,000 people for the US and European countries with available data in 2020 and populations of over one million. Some countries have multiple bar and law societies that represent different types of lawyers (for example, barristers and solicitors in the United Kingdom), which are combined here.

The United States is becoming even more lawyer-intensive. Between 1960 and 1987, the share of GNP going to legal services almost tripled (Sander and Williams 1989). The rate of private litigation per 100,000 people to enforce federal statutes increased from 3 in 1967, to 13 in 1976, 21 in 1986, 29 in 1996, and 40 in 2014 (Burbank and Farhang 2017).

US litigiousness plays an especially significant role in increasing the unpredictability of outcomes because of the politically laden nature of its judicial selection, compared to typically more professional selection abroad. This unpredictability is especially significant because the United States is unusual in granting even lower court judges the ability to halt actions taken by executive branch leaders, such as enjoining infrastructure construction (Kagan 2019; Damaska 1990).

3) Public Sector Planning Capacity. The United States has weak planning capacity (Liscow 2024). For example, Rivadeneira, Dekyi, and Cruz (2023) produce an index on the quality of long-term infrastructure planning, called "long-term strategic vision for infrastructure." Indeed, "national infrastructure frameworks" to enable governments to centralize and direct infrastructure projects are common (McBride, Berman, and Siripurapu 2023). The United States—unlike nearly all the rest of the OECD—did not even have such a plan to analyze.

US federal permitting coordination is very thinly staffed. For example, the Permitting Council, charged with coordinating multi-agency permitting processes—though a significant improvement over no council at all—had only 16 full-time equivalent staff in 2023 (Permitting Council 2024), which the council head emphasizes imposes significant constraints (Siegel 2023). In the aftermath of the Infrastructure Investment and Jobs Act of 2021, which included funding for broadband, the federal government needed facts about internet connectivity across the country to allocate the funds—which took 18 months (*The Economist* 2023).

Most US infrastructure planning happens at the state and local levels. Liscow, Nober, and Slattery (2023) surveyed state department of transportation officials and contractors about their views of personnel at state departments of transportation. Among state officials, 89 percent say that their agencies are understaffed; among contractors, 59 percent agree. Between 1997 and 2021, state employees working on highways (including those working on highway planning) shrunk by 42,300, a 20 percent decline, even as overall state government employment rose. Stronger administrative capacity correlates with lower costs: increasing state-level department of transportation employment per 1,000 residents by one was associated with a 26 percent lower cost per mile of highway.

Understaffing is associated with additional use of outside consultants, and the Liscow, Nober, and Slattery (2023) survey suggests broad agreement (64 percent among government officials, 78 percent among contractors) that this drives up costs—perhaps because consultants are challenging to manage effectively with small in-house staffs, have incentives to push larger projects, and leave states without useful learning-by-doing (Liscow 2024).

In Italy, a country with low transit construction costs, Milan's transit agency has built up so much planning and design capacity that it consults not only on other Italian projects, but also projects abroad (Goldwyn et al. 2023). In contrast, when Boston began building its Green Line mass transit extension, its transit agency had only four to six full-time employees "managing the largest capital project in the agency's history" (Goldwyn et al. 2023, p. 24), leading to poor design choices, reliance on consultants, and high costs. Similarly, "[i]n New York, where consultants largely designed and managed construction for Phase 1 of the Second Avenue Subway, the project management and design contracts were 21 percent of construction costs," whereas in countries with more in-house capacity for planning, like France, "the typical range is 5–10%, with 7–8% most common, and in Italy and Istanbul, it is typically 10%" (Goldwyn et al. 2023, p. 25).

The lack of US planning capacity extends to data. For example, there is no source of consistent or standardized data or measurement of the cost of delayed maintenance (Yousofi and Banta 2023).

The United States may have particularly weak capacity for planning between different levels of government. After passage of the Infrastructure Investment and

Table 2
Two Dimensions of Permitting Regimes

		Capacity of executive to plan: Weak	Capacity of executive to plan: Strong
Power of executive to decide	Weak	United States currently	New Zealand
	Strong	Some developing/middle-income countries (Indonesia, Morocco)	Some of continental Europe (France, Sweden); United States—1950s–1960s interstate highways

Jobs Act of 2021, the federal government asked states to appoint "infrastructure coordinators" to better integrate federal and state planning. A BCG consultant noted, "It goes against a hundred years of how states have worked. . . . It's been hard and awkward for them. But it is a better way to do things" (*The Economist* 2023).

4) Presidential System. The United States is distinguished by a presidential, rather than parliamentary, political system. In a presidential system, different parties may be in control of the legislative and executive branches. In contrast, in a parliamentary system, judicial and administrative decisions can (in principle) reliably be overturned by parliamentary majorities, which lessens the need or desire for external judicial controls because "political controls suffice" (Kagan 2019, p. 256). Furthermore, a presidential system is more likely to have policy "drift" as courts progressively interpret a statute over time, potentially in ways that are not desired by many political coalitions, because often no unified government coalition exists to reverse course.

Taking these factors together, political scientist Robert Kagan (2019) summarizes the unique American way of making law as one of "adversarial legalism," where policy decisions tend to be made by litigation, in a fragmented and decentralized way, with weak executive institutions. In Kagan's telling, the 1960s combined a longstanding mistrust of centralized authority with increasing demands on the US government, in areas from civil rights to environmental protection. The result was a system that relied on the check of citizen litigation through the judiciary, with big penalties at its disposal. At the same time, judges' review of agencies' decision-making "became more searching and demanding" (Kagan 2019, p. 59). Kagan wrote: "It is only a slight oversimplification to say that in the United States, lawyers, legal rights, judges, and lawsuits became functional equivalents for the large central bureaucracies that dominate governance in the activist states of Western Europe" (p. 20). Indeed, political scientist James Q. Wilson (1989, p. 297) contrasts "consensual European administrative practices," in which administrators work informally with other parties to accomplish their goals, to the United States, in which there are strict and lengthy legal procedures that are followed exactly.

In short, "adversarial legalism" is slow, costly, and legally uncertain—and reflects distrust of the state by empowering diffuse parties. It contrasts with other less broadly participatory modes of decision-making, like deference to expert bureaucrats, deference to executive branch political authorities, or negotiation among affected parties. Under adversarial legalism, potential and actual litigants will find it easier to extract concessions from planners, because planners know that they must prevail through multiple levels of executive, legislative, and legal challenges. In this way, the nature of US public administration can lead to the cross-country pattern in the data that we see: higher construction costs and higher civic participation.

Theoretical Framework and Possible Reforms

The international differences across permitting frameworks can be categorized as operating along two axes: executive power and executive planning capacity. The first dimension refers to the amount of discretion granted to the executive branch (federal, state, and local), especially relative to the judiciary. The second dimension is the capacity at all levels of government to plan, which I define as the ability to collect and deploy information, engage the public, and produce cross-project schemes and within-project designs. In particular, notice that the planning dimension is about up-front inputs, while weakness on the executive dimension affects the back end of decision-making, and the extent to which executive branch decisions can be delayed or overridden.

These two axes are presented in Table 2, with examples. The United States today is in the top left, with low executive power and low planning capacity.

In the bottom right are examples with strong executive decisional power and planning capacity. One example is the construction of the US interstate highways, which largely followed a detailed plan developed over about two decades of joint work between the states and the federal government prior to the 1956 passage of the Interstate Highway Act (Federal Highway Administration 2023). Also, the early years of highway-building faced weak judicial or direct public intervention (Lewis 2013), leaving government planners substantial discretion (Wilson 1989). Some continental European countries fit into this quadrant, with strong planning bureaucracies, much-reduced litigation rates, and de facto stronger decision-making powers.

In the bottom left are countries—often developing or middle-income countries—with relatively few judicial (or other) checks on decisions made by the executive and weak planning capacity. For example, Morocco and Indonesia have only weak and deferential judicial review of administrative decision-making (El Abdellaoui 2022; Fenwick 2009). These countries also have weak planning capacity, partly because they rank low on overall state capacity (Hanson and Sigman 2021). In the top right are countries with a well-developed bureaucracy and strong planning capabilities, but weak executive power. New Zealand has very strong statutory compliance requirements that can cause costly delays or litigation (Moore et al. 2021). At the same time, New Zealand has one of the world's highest state capacities on various metrics (Sustainable Governance Indicators 2022).

I now discuss an array of potential US reforms for permitting along each dimension: increasing executive power and improving planning capacity. US permitting practice may be part of a broader US tendency, but that does not mean that reform is impossible. Of course, all reforms have tradeoffs. But after describing possible reforms, I will argue that reforms along both lines can be potentially complementary, in a way that could improve the performance of US permitting.

Power of the Executive to Decide

A variety of reforms would strengthen executive power over permitting. The main benefits of greater executive discretion include reducing delay and costs. The main danger of stronger executive power is greater ability to engage in arbitrary or unlawful activities, arguably the core reason to have judicial review in a system of checks and balances. Indeed, there is a long tradition of trying to use the courts to empower less powerful groups against entrenched and captured bureaucracies (Sabin 2021) and of distrusting centralized authority more generally (Kagan 2019). As well, empowered agencies could be less thorough in their analysis of projects' impacts, potentially leading projects to be more harmful to the human environment.

However, judicial review is an imperfect mechanism for accountability, and it is not the only mechanism. Courts are costly to access, such that well-resourced but nonrepresentative groups could disproportionately benefit from strong judicial review of permitting. In "most democratic countries," an array of other accountability mechanisms include the "administrative supervision and political oversight" that "primarily" holds decision-makers accountable when they are forced to defend decisions in front of political leaders (who wish to be reelected), federal agency funders, legislative appropriation committees, public meeting attendees, and the press (Kagan 2019, p. 249). These tools likely work better than in the 1960s because political organizing around environmental issues has transformed since then (Sabin 2021).

When considering an expansion of executive power, political economy issues arise as well. Greater executive power will increase the likelihood of swings in policy across presidential administrations. Indeed, the public may want to constrain the executive in part from a belief that there is roughly a 50/50 chance that a party with which they strongly disagree will be in charge (Boxell, Gentzkow, and Shapiro 2024).

Here, I list some steps in four broad noncomprehensive areas that could be taken to increase executive power in the specific area of permitting—some appropriate as legislation in the short-term, some more aspirational: (1) shifting legal power away from project opponents; (2) facilitating popular decision-making or effective negotiation between the government and the community; (3) allowing more tailoring of the rules depending on the substance of the environmental or economic impacts; and (4) selective centralization of authority.

In the first area, there are several ways to shift power toward builders and the government over project opponents. One way is that Congress could revise "hard look" review by raising the bar to halt a project. Congress could require that, to nullify a permit, courts have to find that something omitted from the review would likely have impacted the ultimate decision (a so-called "substantial performance" or "materiality" standard). As a result, projects could proceed unless there is a significant oversight. Such a standard resembles German judicial review by administrative courts of environmental assessments (Ladeur and Prelle 2001).

The power to halt projects during litigation could also be limited. Instead, courts could impose financial damages after-the-fact. At present, US courts currently often "enjoin" projects (that is, stop them from proceeding) while litigation over permitting is ongoing, in this way strengthening the power of project opponents. The rules for granting such injunctions could be more restrictive in certain cases (Dourado 2023; Potter, Datta, and Stapp 2022), perhaps those without irreversible harms (like destroying a precious forest or an endangered species, which would in any case be separately covered by the Endangered Species Act). In Germany, a provisional permit may be granted in some cases if the applicant pledges either to provide compensation for damage caused by construction prior to obtaining a nonprovisional license if the decision is positive, or to fund restoration if the license is denied. Thus, a Tesla factory in Germany was able to begin construction with provisional permits, subject to a public objection period and air pollution and water usage review, at risk of facing fines for damages caused by construction (Kolodny and Wayland 2022).

The tradeoffs involved in enjoining projects reflect the classic efficiency-based law-and-economics choice between "liability rules" (damages in this context) and "property rules" (injunctions in this context) (Kraus 2000). High costs of bargaining between the builder and the impacted parties make damages preferable, because these costs prevent efficient bargained solutions—and judicially imposed ex post damages in principle create the right incentives while avoiding the need to bargain. In the permitting context, bargaining costs are often very high because no one can commit all parties to foregoing litigation, so enjoining construction can lead to inefficient delays—including those strategically caused by project opponents. On the other hand, when courts' calculation of harm to impacted parties is challenging, damages work less well: courts' systematic misestimation of damages would produce incorrect incentives for builders. In the permitting context, transaction costs are high and estimating damages is challenging, leaving the theoretical outcome ambiguous. Considerations of equity and procedural justice make clear recommendations even more challenging. Of course, the rule could vary depending upon the ease of bargaining and calculating damages in a given context.

There are other possibilities for limiting the power of project opponents as well, such as restricting standing to sue, automatic approval for projects when review time limits are exceeded, shifting attorneys' fees onto losing parties (or possibly even requiring losing plaintiffs to compensate for delays), expediting judicial review by going directly to appellate courts (thereby weaking the threat of project opponents), or expanding the use of categorical exclusions (Liscow 2025).

A second issue for reform is the challenge facilitating either popular decision-making or effective negotiation between the government and project opponents. Part of the issue is the fundamental commitment problem facing negotiations conducted by the executive: it can strike a deal with some parties to reach a mutually acceptable outcome, but then others can still sue, limiting the incentives to negotiate. One response is to make the negotiations between interested parties more binding. For example, the government could designate interest groups who would then choose representatives, among whom a certain level of agreement would be sufficient for a plan to proceed, with a rule that only a significant failure of the process could lead to subsequent litigation. Something like this approach has at times been employed in the US regulatory context in the form of "negotiated rulemaking," though crucially without any restrictions on litigation. Under the "corporatist" continental European model, representatives of interest groups (often labor unions and industry) bindingly negotiate together with government. Of course, establishing which groups count and who speaks for whom is challenging, especially with historically fragmented representation in the United States (Kagan 2019).

Another way to facilitate negotiation would be to allow more cash incentives to encourage local acquiescence. The fundamental issue is agencies often lack the most efficient tool—cash—to encourage voluntary acceptance of projects. Development sometimes gets expensive because local agreement is achieved using the limited tools available to government agencies, especially building more stuff. For example, in a suburb outside of Detroit, the Federal Highway Administration built large, expensive plazas of questionable value on top of a highway so that a community was less disrupted (Brooks and Liscow 2023). An alternative is providing cash, which could be more valued locally and less expensive than more infrastructure—and more effective at easing opposition. One model for energy is a revenue-sharing agreement with municipalities in exchange for expedited permitting, in a way that bindingly preempts subsequent litigation. Notably, this use of cash payments goes back to the building of the New York City street grid in the early nineteenth century, where taxes and compensation depended on the net damage to each landowner's property caused by a street's construction (Koeppel 2015). Of course, greater availability of cash could also increase costs, because it could weakens governments' ability to commit to not providing compensation, but governments could experiment. In any case, the reform would be most effective if the outcome binds the entire community.

Another possibility for enhancing the commitment of nonjudicial solutions would be for certain legislative decision-making (or referenda) to supersede judicial review. Where a legislative body or voter pool in an election sufficiently encompasses the environmental costs and benefits of a project, it could become a rule that the project is subject to weaker judicial review (and possibly other permitting requirements), on the presumption that the political process has already weighed the costs and benefits. There could still be environmental review, so that decisionmakers are informed. For example, when a city council or local referendum approves a solar or wind farm—and there is not a credible claim of significant environmental damage external to the community (such as harming endangered species), the project could proceed without additional judicial review (absent a gross procedural failing). A similar approach was taken as far back as post–Revolutionary France, in which the decision of a village council on draining communal lands replaced the complex "feudal property rights, local privileges, and easements" that had led to protracted litigation and thwarted agricultural expansion (Kagan 2019, p. 238).

A third way to enhance executive power would be applying to permitting decisions a kind of cost-benefit analysis that tailors requirements to the costs of delay.

When NEPA was passed, it was often the case that building things (like highways) was bad for the environment; with global climate change, often the reverse is the truth now. For example, Congress could adopt a standard by which an agency could designate categories high-priority projects (say, important interstate electricity transmission projects) that would be subject to streamlined environmental and judicial review. Or Congress could designate the categories. This would mean adding more substance based on policy goals to the largely procedural environmental review statutes. For example, New York State has adopted an expedited review process under one central agency for renewable energy projects (Gerrard and McTiernan 2020).

Finally and relatedly, certain powers could be centralized in the federal government. The question of centralization largely boils down to thorny questions of federalism: localism promotes attention to local preferences, but encourages coordination failures across jurisdictions, as well as a failure to reflect externalities beyond local jurisdictions. Notably, since the passage of NEPA, climate change means that the impact of clean energy projects have become less local in scope, making the intensive focus on local impacts less apt. Electricity transmission for clean energy, for example, involves lots of externalities and coordination. To build an interstate natural gas pipeline, the Federal Energy Regulatory Commission (FERC) must approve the project, a decision that preempts local permitting. For non-hydroelectric electricity transmission—for example, transport electricity produced by wind or solar power—FERC does not preempt state and local regulations (with limited exceptions). More expansive FERC preemption would enhance the power of the federal government to approve such transmission lines. An alternative model is Federal Communications Commission permitting for cell towers, which allows local governments to continue regulating cell towers, but prevents them from imposing outright bans and offers builders expedited court proceedings to contest decisions (Dodge 2023). The same principle applies to municipalities within states, reflected in New York's centralization of some decisions around siting of renewable energy projects.

This range of options is provided to start a discussion of concrete possibilities, with the knowledge that each of these options involves potential variations, combinations, and tradeoffs.

Capacity of the Executive to Plan

Better planning capacity can result in the acquisition of more information about both technical details and public preferences, as well as better ways of incorporating that information, yielding better, cheaper, and more-quickly-built infrastructure (Liscow 2024). Infrastructure requires tremendous technical sophistication. Also, market price signals are often not especially helpful for decisions about building and maintaining free-of-charge roads, nor for regulated energy industries, nor for estimation of construction's negative externalities. Government activity, planned or unplanned, coordinated or uncoordinated, will be required one way or another. There is growing empirical evidence in economics on the benefits of strong bureaucratic capacities (Best, Hjort, and Szakonyi 2023; Decarolis et al. 2020; Liscow, Slattery, and Nober 2023).

The direct costs of greater planning capacity are not high, relative to the size of the projects under discussion, as noted earlier for environmental review specifically. The same is true of planning more generally. For example, by the time construction of the interstate highways was nearly complete in 1991, planning, engineering, research, and other miscellaneous expenses still made up less than 10 percent of total costs compared to those dedicated to construction and right of way expenditures (Brooks and Liscow 2023).

But increased capacity that empowers technocrats at the expense of the public, rather than increasing public input, is problematic. Steps intended to improve planning, and to elevate the role of planning, can go astray, since planners could have views that do not represent those of the public. This is especially so because the process of participation itself may matter.

If the focus is on *federal* planning capacity, rather than state and local planning capacity, one can argue that disempowering state and local governments—who may be both more in tune with local constituencies politically and more knowledgeable because they are closer to the ground—could be costly, notwithstanding the potential benefits of greater cross-state coordination.

With these tradeoffs in mind, I list three broad areas of steps to improve planning capacity: (1) streamlined processes; (2) improved and standardized public participation; and (3) more planning staff and better data.

The justification for streamlining processes is not just to improve speed, but also to clarify what evidence needs to be provided and what arguments need to be addressed. The Permitting Council aims to improve permitting through better coordination, establishment of best practices, publicly posted project timelines, and other means, and requires a single document to satisfy NEPA rules, organized by the lead agency where possible. The Fiscal Responsibility Act of 2023 took additional steps. It somewhat loosened certain NEPA requirements. In particular, it clarified the role of a "lead agency" within the review process, set page limits and deadlines for the preparation of environmental reviews, established a process for agencies to use each other's categorical exclusions, and promoted the development of a single environmental review document.

Another way to streamline is to do one environmental review for a large area, rather than project-by-project, reducing duplicative work. Such "programmatic reviews" would speed up permitting, even if a subsequent, but shorter, review is required (Biber and Ruhl 2014). As an example, an environmental impact statement is currently underway for a range of solar power projects in the western United States (US Department of the Interior 2024). In the European Union, member countries now can use programmatic reviews to designate areas promising for renewable energy, typically enabling an exemption from project-level environmental review.[3]

[3] Directive (EU) 2023/2413 of the European Parliament and of the Council of 18 October 2023 amending Directive (EU) 2018/2001, Regulation (EU) 2018/1999 and Directive 98/70/EC as regards the promotion of energy from renewable sources, and repealing Council Directive (EU) 2015/652, 2023 O.J. L. 31.10.

One ultimate goal of streamlining is sometimes called "one-stop shopping." The US model of fragmented power, with independent but potentially overlapping agency jurisdictions (partly to grant each interest group their own government advocate) is liable to duplication of agency procedures and failures of coordination (Kagan 2019). For example, sometimes agencies grant permits in seriatim, rather than simultaneously. Legislation could reduce the fragmentation of decision-making—and encourage simultaneous work.

If permits are to be credible when they come under challenge, then improved and well-structured early-stage participation, both in terms of soliciting information and achieving a stable political settlement, should be a key part of planning (Bozuwa and Mulvaney 2023). Current processes are often not representative because powerful groups tend to dominate, and Rossi and Stack (2023) suggest a number of ways of increasing participation. For example, agencies could develop a "representation floor," in which at the outset they identify groups whose participation should be included and in the final environmental review document say how much those groups actually ended up participating. In addition to ordinary political incentives, the practice of adequately seeking and responding to a broadly participatory set of comments could be enforced through nonjudicial means, including an ombudsman or administrative review.

One example of successful participation is the Miami-Dade Transportation Planning Organization's development of an online tool for local planning officials, which suggests different public outreach strategies tailored to the demographics of the affected communities (Miami-Dade Transportation Planning Organization n.d.). Neighboring Broward County has used this tool to target Facebook advertisements to solicit input on transportation infrastructure spending from non-English speakers in their native language. The county also employed a network of more than 150 partner organizations on the ground to connect with otherwise difficult-to-reach groups, efforts that were further supplemented by focus groups with members of underrepresented communities. Ultimately, two-thirds of participants were those who had never before provided feedback on street planning processes (Broward Metropolitan Planning Organization 2019). This type of participation can also be formalized by citizen assemblies, in which representative samples of the impacted population discuss the issues involved with experts and then vote, and have been used throughout the world (Fishkin 2018).

Greater and more systematic public feedback could slow early project stages, but perhaps only modestly with more capacity to collect it—and this feedback could then make subsequent litigation less likely. As well, a formalized process for expanding participation could (though need not) be combined with a reform that makes decisions binding and preempts further litigation. The need for the cudgel of litigation depends in part on the extent to which one believes that the executive branch—in the absence of judicial review—will conduct good outreach based on ordinary political incentives and bureaucratic practices. At the same time, the threat of litigation resulting from a deviation from settled practices may currently actually be putting a damper on participation by discouraging agencies from "engaging in experimental collaboration efforts" (O'Connor Center 2000, p. 13). Indeed, as noted earlier, the review process is now often oriented around litigation-proofing, rather than actual consultation. Also, whatever the form of participation, experts would still need to play a large informational role given the inherently technical nature of many of the decisions (Liscow and Markovits 2022).

Finally, more staffing and better data would help. The Inflation Reduction Act in 2022 increased funding for permitters. More staffing, especially at the state and local level, would help further. Also important are better data, which could include permitting requirements and projected costs and timelines, as well as data on actual final timelines and project costs, to allow analysis by the government, researchers, and the broader public. Better facts could also reduce costs by reducing the risk premium demanded by private contactors and financers (Loyola 2023). The tradeoff for each of these is the financial cost—and, in the case of data, the need to impose more rules on data provision, which could themselves prove burdensome.

Executive Power and Planning Capacity Together: A Green Bargain?

The question of exactly what form permitting should take raises deep questions about the administrative state, democracy, and the environment: How much should the public trust the executive branch to make decisions, especially in light of ongoing political polarization? How do government institutions best ascertain and aggregate the views of the public? How important is the right of a sustained hearing in court for any party affected by a project?

Theories of permitting reform fall into two types. The first describes a tragic choice with ineluctable tradeoffs: on one side, community participation and environmental quality; and on the other side, speed and efficiency. The second possibility is that evolving circumstances have led the United States to a situation in which, while any reform will face tradeoffs, some of the available reforms may not need to violate major values.

After all, the US permitting regime was historically intended largely as defensive weapon to ensure that citizens' environmental and social concerns were taken into account, when the government seemed to be failing to do so (Ruhl and Salzman 2023). NEPA was passed at a particularly strong moment of government distrust (Sabin 2021). It is not at all obvious that the writers of the law intended the system the United States has now, as decades of layers of "hard-look" judicial doctrine and administrative practice have evolved. There is no reason to think that using courts as the final arbiters of contested permits will arrive at a socially optimal outcome. Courts are good at slowing things down, which is good if building is bad. They can be good at protecting against unlawful behavior and executive branch abuses of power. But a permitting process based on the possibility of years of judicial review also imposes costs, uncertainty, and delay. In some cases, the pro-environment step may be to move ahead with the green energy or the mass transit project, rather than not to build.

I believe that the two dimensions of executive power and planning capacity should be viewed as complements, rather than substitutes; that is, a government with more capacity to plan better on the front end can also be trusted with more discretion to build on the back end. Combining reforms along the two dimensions could lead to a "green bargain," with improvements for efficiency, the environment, and democracy. Enhancing executive power, especially by curtailing litigation, could improve efficiency, increasing construction speed and reducing costs. The resulting increased deployment of renewable energy and transit would improve the environment. Improvements in planning capacity, including ones that promote participation—especially by disadvantaged groups—could make the system more "democratic." After all, broadening public participation and moving it upstream could well be a better way of generating outcomes that reflect public preferences than a series of not-in-my-backward lawsuits brought by small special interest groups. There is little "democratic" about a small handful of people using the courts to hold up projects that have been thoroughly evaluated, with issues widely aired.

Of course, in a situation where people disagree over whether or how a project will be built, no outcome can satisfy everyone. Similarly, no set of reforms can satisfy everyone. The legitimacy of the "green bargain" rises and falls on the structure and mechanisms of participation, as well as its ability to protect environmental and social goals. It runs a risk of slowing down the process in the early stages, even if it is faster and more certain in later states. A standard function of courts is limiting abuses of power by the executive branch; excessive reform could undermine this crucial function of courts. And there are deep currents in US political history—such as a desire for checks-and-balances, a distrust of authority, decentralization, and a powerful judiciary—that could make reform hard. But with all such concerns duly noted, the "green bargain" suggests that in exchange for improved planning and broader early-stage participation, the executive should have greater scope for decision-making.

■ *Thanks to Eli Dourado, Matt Eisenson, Don Elliott, Michael Gerrard, Josh Macey, Dejan Makovsek, Adam Millard-Ball, Susan Rose-Ackerman, J. B. Ruhl, David Schleicher, Alec Stapp, Timothy Taylor, Steve Teles, Heidi Williams, Gui Woolston, the editors, and participants at the Administering a Democratic Political Economy conference and Yale Law School Faculty Workshop for helpful comments. Thanks to Raj Bhargava, Ashlee Fox, Shreyas Mehta, Krister Rasmussen, Safia Sayed, Manav Singh, Nick Whitaker, Connor Wilke, and Taylor Wurts for excellent research assistance.*

References

Adelman, David, and Robert L. Glicksman. 2018. "Presidential Politics and Judicial Review." *Arizona State Law Journal* 50 (3): 3–69.

Allen, James. 2005. "NEPA Alternatives Analysis: The Evolving Exclusion of Remote and Speculative Alternatives." *Journal of Land Resources and Environmental Law* 25: 287.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and Topography of the Spatial Economy." *Quarterly Journal of Economics* 129 (3): 1085–1140.

American Bar Association. 2022. "ABA National Lawyer Population Survey: Historical Trend in Total National Lawyer Population 1878-2022." https://www.americanbar.org/content/dam/aba/administrative/market_research/total-national-lawyer-population-1878-2022.pdf (accessed January 2, 2024).

Bagley, Nicholas. 2021. "The Procedure Fetish." Niskanen Center, December 7. https://www.niskanencenter.org/the-procedure-fetish.

Bennon, Michael, and Devon Wilson. 2023. "NEPA Litigation Over Large Energy and Transport Infrastructure Projects." *Environmental Law Reporter* 53 (10): 10836–59.

Best, Michael, Jonas Hjort, and David Szakonyi. 2023. "Individuals and Organizations as Sources of State Effectiveness." *American Economic Review* 113 (8): 2121–67.

Biber, Eric, and J. B. Ruhl. 2014. "The Permit Power Revisited: The Theory and Practice of Regulatory Permits in the Administrative State." *Duke Law Review* 64 (2): 133–234.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2024. "Cross-Country Trends in Affective Polarization." *Review of Economics and Statistics* 106 (2): 557–65.

Bozuwa, Johanna, and Dustin Mulvaney. 2023. "A Progressive Take on Permitting Reform: Principles and Policies to Unleash a Faster, More Equitable Green Transition." Roosevelt Institute, August 22. https://rooseveltinstitute.org/publications/a-progressive-take-on-permitting-reform/.

Briffault, Richard. 1990. "Our Localism: Part I—The Structure of Local Government Law." *Columbia Law Review* 90 (1): 1–115.

Brinkman, Jeffrey, and Jeffrey Lin. 2024. "Freeway Revolts! The Quality of Life Effects of Highways." *Review of Economics and Statistics* 106 (5): 1268–84.

Brooks, Leah, and Zachary Liscow. 2023. "Infrastructure Costs." *American Economic Journal: Applied Economics* 15 (2): 1–30.

Broward Metropolitan Planning Organization. 2019. "Broward Complete Streets Master Plan." https://www.browardmpo.org/images/Broward_MPO_CSMP_final.pdf.

Burbank, Stephen B., and Sean Farhang. 2017. *Rights and Retrenchment: The Counterrevolution against Federal Litigation*. Cambridge University Press.

Bureau of Transportation Statistics. 2023. "National Transportation Statistics Table 1–10: U.S. Oil and Gas Pipeline Mileage." https://www.bts.gov/content/us-oil-and-gas-pipeline-mileage (accessed October 23, 2023).

Caro, Robert A. 1974. *The Power Broker: Robert Moses and the Fall of New York*. Alfred A. Knopf.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Effects of Neighborhoods on Intergenerational Mobility II: County Level Estimates." *Quarterly Journal of Economics* 133 (3): 1163–1228.

Coleman, James W. 2019. "Pipelines & Power-lines: Building the Energy Transport Future." *Ohio State Law Journal* 80 (2): 263–308.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, et al. 2023. *V-Dem Codebook v13*. V-Dem Institute, University of Gothenburg.

Council of Bars and Law Societies of Europe. 2021. "CCBE Lawyers' Statistics 2020—Full Members." https://www.ccbe.eu/fileadmin/speciality_distribution/public/documents/Statistics/EN_STAT_2020_Number-of-lawyers-in-European-countries.pdf (accessed January 2, 2024).

Council on Environmental Quality. 1997. "The National Environmental Policy Act: A Study of Its Effectiveness after Twenty-Five Years." https://ceq.doe.gov/docs/ceq-publications/nepa25fn.pdf.

Council on Environmental Quality. 2016. "The Fourth Report on Cooperating Agencies in Implementing the Procedural Requirements of the National Environmental Policy Act (NEPA)." https://ceq.doe.gov/docs/ceq-reports/Attachment-A-Fourth-Cooperating-Agency-Report_Oct2016.pdf.

Council on Environmental Quality. 2019. "Length of Environmental Impact Statements (2013–2017)." https://ceq.doe.gov/docs/nepa-practice/CEQ_EIS_Length_Report_2019-7-22.pdf.

Council on Environmental Quality. 2020. "Environmental Impact Statement Timelines (2010–2018)." https://ceq.doe.gov/docs/nepa-practice/CEQ_EIS_Timeline_Report_2020-6-12.pdf.

Council on Environmental Quality. n.d. "NEPA Litigation Surveys: 2001–2013." https://ceq.doe.gov/docs/ceq-reports/nepa-litigation-surveys-2001-2013.pdf.

Currie, Janet, and Reed Walker. 2019. "What Do Economists Have to Say about the Clean Air Act 50 Years after the Establishment of the Environmental Protection Agency?" *Journal of Economic Perspectives* 33 (4): 3–26.

Damaska, Mirjan R. 1990. "Reflections on American Constitutionalism." *American Journal of Comparative Law* 38 (S): 421–443.

Decarolis, Francesco, Leonardo M. Giuffrida, Elisabetta Iossa, Vincenzo Mollisi, and Giancarlo Spagnolo. 2020. "Bureaucratic Competence and Procurement Outcomes." *Journal of Law, Economics, and Organization* 36 (3): 537–97.

de Tocqueville, Alexis. 1835. *Democracy in America*. D. Appleton and Company.

DeWitt, Piet, and Carole A. DeWitt. 2008. "How Long Does it Take to Prepare an Environmental Impact Statement?." *Environmental Practice* 10 (4): 164–74.

Dodge, Joel. 2023. "The '90s Cellphone Law That Could Speed Up the Renewables Rollout." *Heatmap News*, April 12. https://heatmap.news/politics/telecommunications-act-climate-renewable-energy.

Dourado, Eli. 2023. "Bringing NEPA Back to Basics." Center for Growth and Opportunity. https://www.thecgo.org/wp-content/uploads/2023/03/Bringing-NEPA-Back-to-Basics-v1.pdf.

The Economist. 2023. "Spending on Infrastructure has Fallen in Real Terms in America." *The Economist*, November 22. https://www.economist.com/united-states/2023/11/22/spending-on-infrastructure-has-fallen-in-real-terms-in-america.

Einstein, Katherine Levine, Maxwell Palmer, and David M. Glick. 2019. "Who Participates in Local Government? Evidence from Meeting Minutes." *Perspectives on Politics* 17 (1): 28–46.

Eisenson, Matthew, Jacob Elkin, Harmukh Singh, and Noah Schaffir. 2024. *Opposition to Renewable Energy Facilities in the United States: June 2024 Edition*. Sabin Center for Climate Change Law. https://scholarship.law.columbia.edu/sabin_climate_change/226.

El Abdellaoui, Abdelkarim. 2022. "The Right of Individuals to Access Constitutional Justice and Its Role in Protecting Rights and Liberties in Morocco." *Rowaq Arabi* 27 (3): 83–101.

Emerson, Kirk, and Elizabeth Baldwin. 2019. "Effectiveness in NEPA Decision Making: In Search of Evidence and Theory." *Journal of Environmental Policy and Planning* 21 (4): 427–43.

Energy Technologies Area, Berkeley Lab. 2023. "Queued Up: Characteristics of Power Plants Seeking Transmission Interconnection." https://emp.lbl.gov/queues (accessed January 20, 2024.

European Committee of the Regions, Division of Powers. n.d. "France Transport." https://portal.cor.europa.eu/divisionpowers/Pages/France-Transport.aspx (accessed January 6, 2024).

Federal Highway Administration. 2023. "Interstate System." https://www.fhwa.dot.gov/programadmin/interstate.cfm (accessed January 9, 2024).

Federal Highway Administration. 2024. "FOIS Major Projects." https://fhwaapps.fhwa.dot.gov/foismpp (accessed January 31, 2024).

Fenwick, Stewart. 2009. "Administrative Law and Judicial Review in Indonesia: The Search for Accountability." In *Administrative Law and Governance in Asia*, edited by Tom Ginsburg and Albert H. Y. Chen, 329–35. Routledge.

Fishkin, James S. 2018. "Random Assemblies for Lawmaking? Prospects and Limits." *Politics and Society* 46 (3): 359–79.

Flyvbjerg, Bent, and Dan Gardner. 2023. *How Big Things Get Done: The Surprising Factors that Determine the Fate of Every Project, from Home Renovations to Space Exploration and Everything in Between*. Penguin Random House.

Gerrard, Michael B., and Edward McTiernan. 2020. "New York's New Statute on Siting Renewable Energy Facilities." *New York Law Journal* 263 (93).

Goldwyn, Eric, Alon Levy, Elif Ensari, and Marco Chitti. 2023. *Transit Costs Project: Understanding Transit Infrastructure Costs in American Cities*. NYU Marron Institute of Urban Management.

Gordon, Tracy, and David Schleicher. 2015. "High Costs May Explain Crumbling Support for US Infrastructure." Urban Institute, March 31. https://www.urban.org/urban-wire/high-costs-may-explain-crumbling-support-us-infrastructure.

Hanson, Jonathan K., and Rachel Sigman. 2021. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research." *Journal of Politics* 83 (4): 1495–1510.

Hinkle, Madison, and Jesse Richardson. 2022. "How Environmental Litigation Has Turned Pipelines Into Pipe Dreams." *Environmental Law Reporter* 52 (7): 10558–83.

Howard, Philip K. 2015. "Two Years Not Ten Years: Redesigning Infrastructure Approvals." Common Good. https://static1.squarespace.com/static/5db4d0eacb29b173254203d2/t/5fd39d3ac4a2b267e284960c/1607703867904/2YearsNot10Years.pdf.

Jenkins, Jessie D., Jamil Farbes, Ryan Jones, Neha Patankar, and Greg Schivley. 2022. "Electricity Transmission Is Key to Unlock the Full Potential of the Inflation Reduction Act." Unpublished.

Kagan, Robert A. 2019. *Adversarial Legalism: The American Way of Law*. Harvard University Press.

Keiser, David A., and Joseph S. Shapiro. 2019. "Consequences of the Clean Water Act and the Demand for Water Quality." *Quarterly Journal of Economics* 134 (1): 349–96.

Kimmell, Kenneth, and Dawn Stolfi Stalenhoef. 2011. "The Cape Wind Offshore Wind Energy Project:

A Case Study of the Difficult Transition to Renewable Energy." *Golden Gate University Environmental Law Journal* 5 (1): 197–225.

Koeppel, Gerard. 2015. *City on a Grid: How New York Became New York.* De Capo Press.

Kolodny, Lora, and Michael Wayland. 2022. "Tesla Factory in Germany Gains Conditional Approval to Start Commercial Production." *CNBC*, March 4. https://www.cnbc.com/2022/03/04/tesla-factory-in-germany-gains-approval-to-start-commercial-production.html.

Kraus, Michael I. 2000. "Property Rules vs. Liability Rules." *Encyclopedia of Law and Economics* 2: 782–794.

Ladeur, Karl-Heinz, and Prelle, Rebecca. 2001. "Environmental Assessment and Judicial Approaches to Procedural Errors–A European and Comparative Law Analysis." *Journal of Environmental Law* 13 (2): 185–98.

Lewis, Tom. 2013. *Divided Highways: Building the Interstate Highways, Transforming American Life.* Cornell University Press.

Liscow, Zachary. 2024. "State Capacity and Infrastructure Costs." Unpublished.

Liscow, Zachary. 2025. "Reforming Permitting to Build Infrastructure." Unpublished.

Liscow, Zachary, and Daniel Markovits. 2022. "Democratizing Behavioral Economics." *Yale Journal on Regulation* 39 (3): 1217–90.

Liscow, Zachary, William Nober, and Cailin Slattery. 2023. "Procurement and Infrastructure Costs." NBER Working Paper 31705.

Loyola, Mario. 2023. *Global Infrastructure Permitting: A Survey of Best Practices.* Loyola Associates Research and Consulting for the Competitive Enterprise Institute and the Permitting Institute.

Marshall, Tim. 2011. "Planning at the National Level in Europe in Relation to Major Infrastructure." *European Planning Studies* 19 (5): 887–905.

McBride, James, Noah Berman, and Anshu Siripurapu. 2023. "The State of U.S. Infrastructure." Council on Foreign Relations, September 20. https://www.cfr.org/backgrounder/state-us-infrastructure.

Miami-Dade Transportation Planning Organization. n.d. "Community Involvement." http://tpotransportationoutreachplanner.org/mpotop/ (accessed January 10, 2024).

Miles, Thomas J., and Cass R. Sunstein. 2008. "The Real World of Arbitrariness Review." *University of Chicago Law Review* 75 (2): 761–814.

Moore, David, Jeff Loan, Sally Wyatt, Kelvin Woock, Sally Carrick, and Zabard Hartmann. 2021. "The Cost of Consenting Infrastructure Projects in New Zealand." Unpublished.

Nicholson, Charles P., ed. 2022. 2022 *Annual NEPA Report.* National Association of Environmental Professionals.

NRDC. 2021. "Keystone XL Pipeline Lawsuits." https://www.nrdc.org/court-battles/keystone-xl-pipeline (accessed January 15, 2024).

NYU Transit Costs Project. 2020. "Data." Marron Institute, New York. https://transitcosts.com/data/ (accessed February 2, 2024).

NYU Transit Costs Project. 2022. "What the Data Is Telling Us." https://transitcosts.com/new-data/ (accessed January 6, 2024).

O'Connor Center for the Rocky Mountain West (CRMW) and Institute for Environment and National Resources (IENR). 2000. *Reclaiming NEPA's Potential: Can Collaborative Processes Improve Environmental Decision Making?* CRMW and IENR.

Pahlka, Jennifer. 2023. *Recoding America: Why Government Is Failing in the Digital Age and How We Can Do Better.* Metropolitan Books.

Pepper, Elly. 2015. "Never Eliminate Public Advice: NEPA Success Stories." Natural Resources Defense Council, February 1. https://www.nrdc.org/resources/never-eliminate-public-advice-nepa-success-stories.

Permitting Council. 2024. *Fiscal Year 2025 Budget Request.* US Department of Transportation.

Potter, Brian, Arnab Datta, and Alec Stapp. 2022. "How to Stop Environmental Review from Harming the Environment." Institute for Progress, September 13. https://ifp.org/environmental-review/.

Ramseyer, Mark J., and Eric B. Rasmusen. 2010. "Comparative Litigation Rates." John M. Olin Center for Law, Economics, and Business, Harvard Law School Discussion Paper 681.

Rivadeneira, Ana Maria Ruiz, Tenzin Dekyi, and Lorena Cruz. 2023. "OECD Infrastructure Governance Indicators: Conceptual Framework, Design, Methodology and Preliminary Results." OECD Working Papers on Public Governance 59.

Rossi, Jim, and Kevin M. Stack. 2023. "Representative Rulemaking." *Iowa Law Review* 109 (1): 1–50.

Ruhl, J. B., and James Salzman. 2023. "The Greens' Dilemma: Building Tomorrow's Climate Infrastructure Today." *Emory Law Journal* 73 (1): 1–82.

Sabin, Paul. 2021. *Public Citizens: The Attack on Big Government and the Remaking of American Liberalism.*

W. W. Norton and Company.

Sander, Richard H., and E. Douglass Williams. 1989. "Why Are There So Many Lawyers? Perspectives on a Turbulent Market." *Law and Social Inquiry* 14 (3): 431–79.

Schmalensee, Richard, and Robert N. Stavins. 2019. "Policy Evolution under the Clean Air Act." *Journal of Economic Perspectives* 33 (4): 27–50.

Schumacher, Kim. 2019. "Approval Procedures for Large-Scale Renewable Energy Installations: Comparison of National Legal Frameworks in Japan, New Zealand, the EU and the US." *Energy Policy* 129: 139–52.

Seelye, Katharine Q. 2017. "After 16 Years, Hopes for Cape Cod Wind Farm Float Away." *New York Times*, December 19. https://www.nytimes.com/2017/12/19/us/offshore-cape-wind-farm.html.

Siegel, Josh. 2023. "The Biden Administration's New Permitting Fixer." *Politico*, June 28. https://www.politico.com/newsletters/the-long-game/2023/06/28/the-biden-administrations-new-permitting-fixer-00103990.

Solar Energy Industries Association. n.d. "Development Timeline for Utility-Scale Solar Power Plant." https://www.seia.org/research-resources/development-timeline-utility-scale-solar-power-plant (accessed January 5th, 2024).

Sustainable Governance Indicators. 2022. "Exectutive Capacity #4." https://www.sgi-network.org/2022/New_Zealand.

US Bureau of Labor Statistics. 2023. "Occupational Employment and Wages, May 2023." https://www.bls.gov/oes/current/oes193051.htm.

US Department of the Interior. 2024. "Bureau of Land Management Releases Proposed Western Solar Plan." Bureau of Land Management, August 29. https://www.blm.gov/press-release/bureau-land-management-releases-proposed-western-solar-plan.

US Energy Information Administration. 2023. "Electricity Explained." https://www.eia.gov/energyexplained/electricity/electricity-in-the-us-generation-capacity-and-sales.php.

US Government Accountability Office. 2014. *National Environmental Policy Act: Little Information Exists on NEPA Analyses*. GAO-14-370.

Wagner, Wendy, William West, Thomas McGarity, and Lisa Peters. 2021. "Deliberative Rulemaking: An Empirical Study of Participation in Three Agency Programs." *Administrative Law Review* 73 (3): 609–87.

Wilson, James Q. 1989. *Bureaucracy*. Basic Books.

Wolf, M. J., J. W. Emerson, D. C. Esty, A. de Sherbinin, Z. A. Wendling, et al. 2022. *2022 Environmental Performance Index*. Yale Center for Environmental Law and Policy, New Haven, CT. https://doi.org/10.7927/dwt2-9k25 (accessed October 17, 2023).

World Bank. 2023. "World Development Indicators: Population, total." https://databank.worldbank.org/reports.aspx?source=2&series=SP.POP.TOTL&country= (accessed January 2, 2024).

Wyoming Renewable Energy Coordination Committee. 2022. *Guide to Permitting Wind Energy Projects in Wyoming*. State of Wyoming

Yousofi, Fatima, and Susan Banta. 2023. "State and Local Governments Face Persistent Infrastructure Investment Challenges." *Pew Trusts*, February 3. https://www.pewtrusts.org/en/research-and-analysis/articles/2023/02/03/state-and-local-governments-face-persistent-infrastructure-investment-challenges.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 181–204

A Practical Guide to Shift-Share Instruments

Kirill Borusyak, Peter Hull, and Xavier Jaravel

Many economic studies consider units that are exposed differently to a common set of shocks. Consider, for example, the influential Autor, Dorn, and Hanson (2013) study of how the surge in Chinese imports in the 1990s and 2000s affected US local labor markets. They measure regional exposure to this "China shock" by the extent to which workers were employed in industries that saw growing competition with China. This idea is captured by a *shift-share* explanatory variable: the average of national industry-level shifts in US imports from China, weighted by the regional shares of employment across industries. They further construct an *instrumental variable* with a similar shift-share structure: the average of industry growth in Chinese imports among non-US countries, again weighted by industry employment shares of US commuting zones. By using this instrument in a two-stage least squares regression, the authors hope to address potential endogeneity concerns: namely, that US imports from China may be affected by US-specific productivity and demand shocks.

Instruments like these, which sum a common set of shifts with heterogeneous exposure share weights, are often used in studies of labor, trade, macroeconomics, public economics, finance, and more. While such instruments date back at least to Freeman (1975, 1980), the number of papers using them has grown markedly over

■ *Kirill Borusyak is Assistant Professor of Agricultural and Resource Economics, University of California Berkeley, Berkeley, California. Peter Hull is Professor of Economics, Brown University, Providence, Rhode Island. Xavier Jaravel is Professor of Economics, London School of Economics, London, United Kingdom. Their email addresses are k.borusyak@berkeley.edu, peter_hull@brown.edu, and x.jaravel@lse.ac.uk.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20231370.

the last ten years (Goldsmith-Pinkham 2024). Today, around one-eighth of all instruments featured in NBER working papers are explicitly described as shift-share, while many others implicitly have a shift-share structure.

When do such instruments successfully solve endogeneity concerns, and when might they fail? This question is challenging to answer because shift-share instruments leverage two distinct sources of variation, and it is not obvious what properties of each are important. Intuitively, one might view the shifts as helpful because they represent potentially exogenous changes to the system under study. However, these shifts vary at a different level (for example, industries) than the unit of analysis (for example, local labor markets). Are they still useful then? In contrast, the shares vary across units but are usually predetermined (for example, employment shares are measured in a pre-period). So how should their potential exogeneity be understood?

This article gives conceptual answers to these questions and provides practical guidance for using shift-share instruments or assessing the credibility of such instruments when used by others. We build on a recent econometric literature which suggests two distinct paths to identification. One path, developed by Borusyak, Hull, and Jaravel (2022) and Adão, Kolesár, and Morales (2019), leverages many exogenous shifts while making no assumption on the exogeneity of the shares. The second path, proposed by Goldsmith-Pinkham, Sorkin, and Swift (2020), instead focuses on share exogeneity. Each of these two approaches has distinct practical implications regarding appropriate estimators, ways to conduct valid inference, and diagnostic tests.

We begin with a discussion of broad motivations for using shift-share instruments and an overview of the core logic for both paths. We discuss how identification “from the shifts” can be understood as leveraging a shift-level natural experiment, while identification “from the shares” can be viewed as pooling together multiple difference-in-differences designs leveraging heterogeneous shock exposure. We then provide two checklists researchers can follow when implementing a shift-share design, considering the exogenous shifts and exogenous shares approaches in turn. We take an applied perspective throughout, illustrating key concepts and practical steps with examples; see Borusyak, Hull, and Jaravel (2024) for a more technical review.

Online Appendix A answers further practical questions that often arise with shift-share instruments. For instance, we discuss how to interpret estimates as local averages of heterogeneous effects, how to handle multiple instruments and interaction terms, how to approach shift-share instruments where the shifts are measured in-sample (as in Bartik 1991; Card 2009), and whether a leave-out construction of the shifts is helpful in those cases.

Shift-Share Basics

What Are Shift-Share Instruments and Where Do They Come From?

Table 1 lists some prominent examples of shift-share instruments from a variety of settings. We discuss some of these examples in depth below. Here, the table is meant to illustrate some common features of a shift-share research design.

Each study seeks to estimate a causal or structural relationship between two variables measured across a set of units i. The *outcome* variable is denoted y_i. Borrowing standard language from the world of causal inference, we refer to the explanatory variable x_i as the *treatment*. For example, Autor, Dorn, and Hanson (2013) seek to estimate the causal effect of growing exposure to Chinese imports x_i on the growth in local manufacturing employment y_i (among other outcomes) across US regions i. The table shows many other examples of outcomes and treatments: across regions, firms, products, and individuals.

To formalize the goal in such settings, consider a model of the form:

$$y_i = \beta x_i + \gamma' \mathbf{w}_i + \varepsilon_i, \tag{1}$$

where $\mathbf{w}_i$ denotes some vector of observed control variables. Here β is the parameter of interest, capturing the effect of the treatment on the outcome (which for simplicity is assumed to be the same across units). The error term ε_i captures all unobserved determinants of the outcome. We assume throughout that this outcome equation is correctly specified, focusing on consistent estimation of β rather than choosing and interpreting the specification.

Importantly, in writing equation (1), we allow for the possibility of treatment *endogeneity*: that is, a non-zero correlation between x_i and ε_i. In the Autor, Dorn, and Hanson (2013) example this allows US regions with more exposure to Chinese imports to have different unobserved labor market conditions, which would have led to lower or higher manufacturing employment growth in the absence of the China shock. Such endogeneity introduces bias in ordinary least squares estimates of equation (1). A standard solution to this challenge is to find an *instrument* z_i that is plausibly uncorrelated with the unobserved model error ε_i while nevertheless correlated with the endogenous treatment x_i. The parameter β can then be estimated by two-stage least squares.

The instruments in Table 1 are distinguished by their shift-share structure:

$$z_i = \sum_{k=1}^{K} \underbrace{s_{ik}}_{\text{Share}} \underbrace{g_k}_{\text{Shift}}, \tag{2}$$

where $(g_1, \ldots, g_K)$ is a set of shifts that is common to all units and the $(s_{i1}, \ldots, s_{iK})$ are sets of exposure shares that vary across units. In many applications, the shares sum to one for each observation such that z_i is a share-weighted average of the shifts.

In most of the Table 1 examples, the shifts are defined at a different level k than the units i. For example, Bartik (1991) and Autor, Dorn, and Hanson (2013) work with regional outcomes and industry-level shifts. Exceptions are studies of network spillovers where k indexes friends or neighbors of individuals or regions. It is also worth noting that while most examples in Table 1 use a shift-share instrument to address endogeneity in a treatment x_i, some (indicated by an asterisk in the

Table 1
Shift-Share Instrument Examples

Study	Unit (i)	Outcome (y_i)	Treatment (x_i)	Level of shift variation (k)	Instrument (z_i): Share (s_{ik})	Instrument (z_i): Shift (g_k)
Bartik (1991)	Region	Δ Local wage	Δ Local employment	Industry	$\text{Employment}_{ik}/\text{Employment}_i$	National growth of industry employment
Miguel and Kremer (2004)	Individual	Measures of health or education	Number of neighbors selected for deworming*	Individual	$\mathbf{1}\{k \text{ is friend of } i\}$	Dummy of deworming treatment
Card (2009)	Region	Relative wage of migrants vs. natives	Relative employment of migrants vs. natives	Origin country	$\text{Migrant stock}_{ik}/\text{Population}_i$	$\text{New migrants}_k/\text{Migrant stock}_k$
Autor, Dorn, and Hanson (2013)	Region	Δ Local manufacturing employment	Δ Local exposure to Chinese imports	Industry	$\text{Employment}_{ik}/\text{Employment}_i$	Δ Imports from China in other countries
Hummels et al. (2014)	Worker	Wage	Imports of intermediate goods by employer	Product-by-country	$\text{Imports}_{ik}/\text{Imports}_i$	Imports from k to other countries
Nunn and Qian (2014)	Country-by-year	Conflict	Quantity of food aid (wheat) from the US	Year	Fraction of years with non-zero food aid	US wheat production in previous year
Cai, Janvry, and Sadoulet (2015)	Individual	Takeup of insurance	% of friends selected for an information session*	Individual	$\mathbf{1}\{k \text{ is friend of } i\}/\#$ of friends i has	Dummy of information session
Jaravel (2019)	Product category	Inflation and innovation	Δ Quantity demanded	Sociodemographic group	Sales of i to group k/Total sales of i	Population change
Greenstone, Mas, and Nguyen (2020)	Region	Δ Employment	Δ Credit	Bank	Credit market share of k	Estimated credit supply shock
Aghion et al. (2022)	Firm	Δ Firm employment	Δ Firm stock of automation technologies	Technology-by-country	$\text{Imports}_{ik}/\text{Imports}_i$	Δ Imports from k to other countries
Xu (2022)	Region	Δ Exports	Exposure to banking crisis*	Bank	Credit market share of k	Bankruptcy during banking crisis
Franklin et al. (2024)	Local labor market	Wage	Shift-share exposure to the intervention*	Residential neighborhood	$\text{Commuters}_{ik}/\text{Employment}_i$	Dummy of public works intervention
Mohnen (2025)	Region	Δ Young labor market outcome	Retirement rate	Age group (within 45+)	$\text{Population}_{ik}/\text{Population } 45+_i$	National retirement rate at age k

Note: We simplify many of the settings, suppressing the time dimension (except where it is central to the design), controls and fixed effects, interaction terms, log and other transformations of the outcome and treatment, and so forth. Asterisks (*) indicate ordinary least squares regressions, in which the treatment itself is the shift-share with shares s_{ik} and shifts g_k.

treatment column) consider "reduced-form" regressions on z_i itself. We capture this by defining $x_i = z_i$ in such settings.

Researchers might motivate shift-share instruments in different ways. A common motivation arises when the treatment measures the growth of some economic variable over time, and can be decomposed into some start-of-period shares and over-time shifts. Suppose, for example, that $x_i = (X_{i1} - X_{i0})/X_{i0}$ is the growth in employment X_{it} for local labor market i over two periods, $t = 0, 1$. Regional employment can be decomposed across industries: $X_{it} = \sum_k X_{ikt}$ where X_{ikt} denotes the period-t employment of industry k in local labor market i. This leads to a decomposition of regional employment growth rates in terms of period-0 industry employment shares and local industry growth rate shifts:

$$x_i = \sum_k \underbrace{\frac{X_{ik0}}{X_{i0}}}_{\text{Share}} \cdot \underbrace{x_{ik}}_{\text{Local shift}}, \quad \text{for } x_{ik} = \frac{X_{ik1} - X_{ik0}}{X_{ik0}}. \tag{3}$$

A researcher might then construct an instrument by choosing a set of common shifts g_k to replace the local shifts. The shares from the decomposition could be kept or also replaced, for example, with further lagged shares. Instruments constructed in this way tend to be highly correlated with the treatment.

To illustrate this motivation, consider an example inspired by the canonical shift-share instrument of Bartik (1991). The goal is to estimate the inverse elasticity of regional labor supply β relating wage growth y_i to employment growth x_i across regions i. As usual, to estimate a supply elasticity we need an instrument that shifts labor demand. Decomposition (3) captures the idea that x_i averages local employment growth across different industries, x_{ik}, using initial employment shares $s_{ik} = X_{ik0}/X_{i0}$ as weights. The local shifts reflect changes to both labor demand and labor supply. To isolate demand variation, we can form an instrument that keeps the local industry employment shares from the decomposition but introduces a set of common shifts. The shifts are meant to be predictive of the local industry growth rates while only capturing demand variation. Bartik (1991) defines g_k as national industry growth rates, proxying for aggregate demand shifts. One might also define g_k as specific industry demand shifts, such as a change in government subsidies.

Decomposition (3) is helpful for illustrating why the shares in the definition of z_i often, but not always, sum to one for each observation. In the previous example, regional employment shares mechanically sum to one across industries. However, sometimes the instrument is constructed from shifts that could only happen in a subset of industries: say, within the manufacturing sector. Only those industries would appear in the shift-share instrument formula, and the shares would add up to a number smaller than one. We discuss the importance of this below.

Decomposing the treatment is not the only way to arrive at a shift-share instrument. Another common way is by "apportioning" some national changes to units. Online Appendix A.5 illustrates this approach and shows how it relates to the decomposition above. In still other cases, an instrument naturally takes a shift-share form. For instance, many reduced-form studies of how shocks propagate across a network (for example, Cai, Janvry, and Sadoulet 2015) use the fraction of unit i's friends or neighbors who have been selected for some intervention. This variable inherently has a shift-share structure: $z_i = \sum_k s_{ik} g_k$ where g_k is a dummy variable indicating that k has been selected and s_{ik} is a dummy variable indicating that k is a friend of i, scaled by the number of friends i has.

Regardless of the motivation, the core challenge in using such z_i is to argue convincingly that it is *exogenous*; that is, uncorrelated with the model unobservable ε_i. Such arguments are typically made from contextual knowledge about the source of variation in the instrument. The unique challenge with shift-share instruments is that there are two distinct sources of variation: the shifts and the shares. Thus, to argue convincingly that these instruments are exogenous, one must explain what properties of the shifts and shares make z_i uncorrelated with ε_i (rather than simply stating the basic exogeneity restriction of $\text{cov}[z_i, \varepsilon_i] = 0$). We next introduce the two paths for making such arguments.

What Is Identification from Many Exogenous Shifts?

One strategy to ensure that the shift-share instrument z_i is exogenous is to have exogenous g_k. For example, imagine a lottery that randomly assigns a subsidy level g_k to each industry k. In the above labor supply setting, local employment growth x_i can be instrumented for by a weighted average of the subsidies, using initial local employment shares as weights. Subsidies can be viewed as only affecting wages by shifting labor demand and do not have direct effects on labor supply. In general, exogenous shifts should be as-if randomly assigned and should only affect the outcome through the treatment (an *exclusion restriction*).

Shift-based identification stems from a simple observation: a share-weighted average of random shifts is itself as-good-as-random. This is true even if the shares are econometrically endogenous, in the sense that units with different shares may have systematically different unobservables. For instance, regions that specialize in high-skill intensive industries may experience more immigration from certain countries, such that the total employment share of high-skill intensive industries positively correlates with unobserved immigration shocks in the error term. But as long as subsidies are assigned at random across many high- and low-skill intensive industries, on average the places specializing in high-skill intensive industries will have typical values of the instrument. Thus, a shift-share research design based on experimental shifts requires no assumptions on the exogeneity of exposure shares.

While a lottery provides intuition for an idealized experiment, the necessary and sufficient condition for instrument exogeneity is a weaker condition on the shifts: g_k should be uncorrelated with an average of ε_i taken across units with weights s_{ik}. In our running example, this would mean that subsidies g_k—even if not truly randomized—are not systematically higher or lower in industries which are concentrated (in terms of employment shares s_{ik}) in regions with high versus low labor supply shocks ε_i. Violations of this condition are the key threat to identification in the exogenous shifts approach.

Another way to understand the exogenous shifts approach is to view the shift-share instrument as a "translation device" for a set of as-good-as-random shifts to a different level of analysis. For instance, when industry subsidies are as-good-as-randomly assigned, one could imagine running an industry-level analysis that uses the subsidy g_k directly as an instrument for industry employment. Specifying the equation at the level of local labor markets may define a more interesting economic parameter, capturing spillovers when workers move across industries in response to the subsidies. However, the key identification assumption is the same, with the shift-share instrument translating the industry-level natural experiment to local labor markets.

The "weighted average of lotteries" logic highlights two other requirements of the exogenous shifts approach. First, it requires many shifts $g_1, \ldots, g_K$. Otherwise, if K is a small number, the shifts may by chance be correlated with unobservables even if they are truly random.[1] This can be viewed as an instance where the law of large numbers does not apply: there are effectively only a few exogenous comparisons, regardless of how many units are observed.

Second, the shares have to add up to one such that the shift-share instrument has an interpretation as a share-weighted *average* of shifts rather than a share-weighted *sum*. Otherwise, even if shifts are drawn fully at random, the instrument may systematically vary across units through the sum of shares. We discuss below how the exogenous shifts approach extends in this "incomplete shares" case.

What Is Identification from Exogenous Shares?

A different strategy to ensure shift-share exogeneity is to have exogenous shares. What does this mean exactly? One could imagine the set of s_{ik} being as-good-as-randomly assigned to units, as if drawn in a lottery, and satisfying an exclusion restriction (that the shares affect the outcome only via the treatment of interest). Alternatively, when the outcome is measured in changes, one may interpret share exogeneity as a set of parallel trends conditions similar to ones used in difference-in-differences strategies. That is, for s_{ik} to be uncorrelated with ε_i one could assert that—if not for any change in the treatment—outcomes would have trended similarly for units that were more versus less exposed to k. Shares are exogenous when such parallel trends conditions hold for each k.

To make this logic concrete, consider an example inspired by Card (2009) who estimates the (inverse) elasticity of substitution between migrant versus native workers in labor demand, β. Here the model (1) relates changes in the relative wages of migrants versus natives between two periods, y_i, to changes in the relative employment of these groups, x_i, across local labor markets. Suppose that between these periods we saw a sudden change in national migrant inflows from a particular origin country κ, such as the sudden inflow of Cuban immigrants following the

[1] We note that it does not matter whether the g_k take many distinct values. For instance, assigning a 10 percent subsidy to some of the many industries (and 0 percent to the rest) should be viewed as having many shifts.

Mariel Boatlift studied by Card (1990). One might be willing to assume that regions that were more or less exposed to this inflow, as captured by the initial share of migrants from Cuba $s_{i\kappa}$, would have seen similar trends in labor demand for migrant versus native labor: that is, that $\text{cov}[\varepsilon_i, s_{i\kappa}] = 0$. In this case, the Cuban migrant share would be a valid instrument for identifying β.

Under such share exogeneity, shift-share instruments can be viewed as combining multiple valid share instruments—each operating under the same difference-in-differences logic, but capturing different exposure variation.[2] Indeed, Goldsmith-Pinkham, Sorkin, and Swift (2020) show that shift-share estimates can generally be viewed as pooling together K "one-at-a-time" estimates each using a single s_{ik} share as the instrument. In the above example, this would mean sudden changes in migrant inflows across many origin countries, to different extents. In this case, if a parallel trends condition holds with respect to each exposure share, a shift-share instrument combining them with g_k weights will also be a valid instrument.

Thus, the exogenous shares approach is appropriate when a researcher is comfortable using any of the individual shares as an exogenous instrument. The plausibility of share exogeneity depends on whether there are conceivably any unobserved shocks that affect the outcome via the same (or similar) shares as the ones used to construct the instrument. Even if shares are drawn at random from a lottery, the presence of any such shocks would always lead to parallel trend violations.

The plausibility of share exogeneity is bolstered by constructing the instrument with shares that are "tailored" to the treatment of interest, in the sense of mediating only the shocks to x_i and not a broad set of shocks that might affect y_i. For example, in the literature on the effects of immigration (for example, Card 2009), exposure shares are tailored to the research question: they measure local migration from various origins in the past. This scenario can be contrasted with popular shift-share designs with shares reflecting local industrial composition while studying the regional impacts of specific industry shifts, such as import competition with China in Autor, Dorn, and Hanson (2013) or robotization in Acemoglu and Restrepo (2020). The industry employment shares are "generic," in that they could potentially measure an observation's exposure to other shocks (essentially, any industry shock), many of them unobserved. In studies using such shares, it would not be plausible to make a case for identification based on the exogeneity of shares. Under the exogenous shares view, Autor, Dorn, and Hanson (2013) and Acemoglu and Restrepo (2020) use essentially the same instruments (lagged employment shares) for different treatments.

The role of the shifts is secondary with the exogenous shares strategy: Goldsmith-Pinkham, Sorkin, and Swift (2020) show that the shifts affect the weights in their representation of shift-share as pooled one-at-a-time share-instrument

[2] To see this link, note that we can formalize the above example as a setting with only one non-zero immigration shift: that is, $g_\kappa \neq 0$ and $g_k = 0$ for all other host countries $k \neq \kappa$. The resulting shift-share instrument $z_i = \sum_{k \neq \kappa} s_{ik} \cdot 0 + s_{i\kappa} g_\kappa = s_{i\kappa} g_\kappa$ is perfectly collinear with $s_{i\kappa}$, so using this single exposure share as the instrument will produce numerically the same estimate.

estimates, but they do not affect the identification of β so long as the shares are exogenous. The choice of g_k, however, may affect the power of the shift-share instrument. Intuitively, the decomposition (3) suggests that a powerful instrument might use as the g_k the average of shifts x_{ik} across units (for example, replacing the local growth rates of industry employment x_{ik} with the national ones g_k).

Many Exogenous Shifts in Practice

We now describe a list of practical steps for applying shift-share designs with many exogenous shifts. This checklist can also be instructive for assessing the design of existing papers using shift-share instruments. Column 1 of Table 2 summarizes some of the main practical takeaways discussed in this section. We illustrate the checklist in the labor supply setting from above, where g_k represents as-good-as-randomly assigned federal subsidies to industries k. At the end of this section, we discuss several real-world examples.

A Checklist for the Shift-Based Approach

1. Motivate the shift-share strategy with an idealized shift-level experiment. Any compelling instrumental-variable design begins with thinking about what endogeneity bias is being addressed: that is, exactly which unobserved variables (or *confounders*) are likely to bias simple ordinary least squares estimation. For example, when attempting to estimate a labor supply equation with data on local employment growth x_i and local wage growth y_i, the model error ε_i will include unobserved local labor supply shocks (for example, immigration of foreign workers to each region). Because equilibrium employment growth arises from both labor supply and labor demand shocks, it is generally correlated with ε_i, generating bias in ordinary least squares estimates. To estimate β, we need an instrument that is uncorrelated with local labor supply changes.

Once potential confounders are specified, the researcher can describe a hypothetical shift-level experiment which would generate shifts that are unrelated to these sources of bias while nevertheless generating variation in the treatment. For example, one can imagine assigning new federal subsidies at random across industries. Industries receiving larger subsidies are likely to expand their production and thus their demand for local workers, increasing local employment x_i. By virtue of random assignment, these subsidy shifts are unrelated to local labor supply conditions. The experimental ideal is thus useful to clarify exactly the type of shift-level variation one would want for identification.

2. Bridge the gap between the observed and ideal shifts. The next step is to describe how the actual shift-share design used for the empirical analysis approximates the idealized experiment. This may involve (1) specifying some control variables and (2) describing how observed shifts proxy for the ideal ones.

Table 2
Summary of Main Practical Takeaways

	Approach	
	Many exogenous shifts (1)	Exogenous shares (2)
Identification argument	Shifts are as-good-as-randomly assigned and only affect the outcome through the treatment	Each share satisfies parallel trends: the outcomes of units with high versus low shares would have trended the same if not for the treatment
Estimation	Control for the sum of shares (if not one) and shift-share aggregates of any shift-level controls	Check robustness to using share instruments directly: for example, one share at a time or pooled via two-stage least squares or limited information maximum likelihood
Statistical inference	Get exposure-robust standard errors from the equivalent shift-level instrumental variable regression	Use conventional heteroskedasticity- or cluster-robust standard errors
Balance tests	For both the shift-share instrument and the shifts	For both the shift-share instrument and the shares with high Rotemberg weights
Do not use when…	You would not want to use the shifts directly as an instrument in a shift-level regression, for example, because they are too few or endogenous	You would not want to use shares directly as instruments, for example, because they are "generic" (capturing the unit's exposure to many types of shocks)

In our running example, imagine changes in subsidies g_k are not randomized across industries and could provide shift-level variation analogous to the randomized subsidies only conditional on some controls. There could be two types of such controls, depending on whether shift-level or unit-level confounders motivate including them. For the former, one may consider shift-level observables q_k that both correlate with the g_k and can have a direct impact on the outcome of interest. For example, one might worry that subsidies are systematically larger in skill-intensive industries and that immigration from skill-abundant countries shifts labor supply in regions where those industries concentrate. In this case, one would like to control for the indicator of skill-intensive industries in the shift-share specification. But how can this be done if such q_k vary at the industry level while the specification is estimated at the regional level? Borusyak, Hull, and Jaravel (2022) show that the answer is to control for $\sum_k s_{ik} q_k$: shift-share aggregates of the industry-level confounders, with the same exposure shares as in the construction of the instrument. In the skill intensity example, this amounts to controlling for the total regional employment share of all skill-intensive industries. With this control variable, the shift-share instrument will only leverage the variation in g_k which is uncorrelated with the q_k; for example, residual variation in subsidies after controlling for skill intensity.

Controls of the second type arise from unit-level observables which are thought to correlate both with the error term ε_i and with z_i. For example, one might expect labor markets in the US "Rust Belt" to experience different unobserved local labor supply shocks versus other parts of the country, and that industries more concentrated in these states see systematically different subsidies. In this case, a straightforward solution is to control for a Rust Belt indicator.[3]

Even after including the controls, the shifts may only be viewed as proxies for idealized ones. In Autor, Dorn, and Hanson (2013), for example, industry-level productivity shifts in China are unobserved but proxied with the growth of imports from China in non-US countries. In Bartik (1991), labor demand shifts are proxied with national employment growth rates. In those cases, the applicability of the exogenous shifts approach depends on whether the gap between the proxy and ideal shift could be contaminated by confounders. In online Appendix A.11, we show that this problem arises in Bartik (1991).

3. Include the "incomplete share" control. In shift-share designs where the exposure shares s_{ik} do not add up to one—what Borusyak, Hull, and Jaravel (2022) call the "incomplete shares" case—a special control must be included: the sum of shares, $S_i = \sum_k s_{ik}$. To build intuition, recall that with "complete" shares (when $S_i = 1$), the shift-share instrument is a weighted *average* of the shifts so if shifts arise from a pure lottery then the shift-share instrument is also like a lottery outcome. This logic breaks down with incomplete shares, when z_i is a weighted *sum* of the shifts. Then, even with randomly assigned shifts—which have, say, a positive mean—an observation with a higher S_i would systematically get higher values of the instrument. The instrument is thus correlated with the sum of shares, which can in turn be correlated with the error, leading to bias. Controlling for S_i removes the problem, because units with the same S_i get different values of the shift-share only for random reasons.[4]

4. Lag shares to the beginning of the natural experiment. When constructing the shift-share instrument, one needs to decide when to measure the shares. Decomposition (3) suggests measuring them at the beginning of the period of interest, but it is common in practice to lag them further. Is this practice justified?

In the exogenous shifts approach, it is best to measure the shifts at the beginning of the natural experiment that generates them. This avoids the situation where the shifts affect the shares, potentially generating bias.[5] At the same time, shares

[3] An alternative way would be to consider industry-level controls as described above; for example, the share of Rust Belt regions in the industry employment. The relative merits of these two approaches remain unexplored.

[4] In online Appendix A.4, we explain why this solution is typically better than renormalizing the shares to add up to one. We also note that sometimes researchers introduce the shares that add up to one across *observations* instead of shifts. In such a case, it is not appropriate to control for S_i. Instead, the shift-share instrument should be rewritten in a different way consistent with (2); see online Appendix A.5.

[5] Not every response of the shares to past shifts makes the shift-share instrument endogenous: if this response is not related to the error terms, there is no problem. But it is possible to imagine situations where the bias would arise. Following footnote 32 in Borusyak, Hull, and Jaravel (2022), consider the labor supply setting and imagine that subsidies now occur in two periods. Supposing regions vary in labor market flexibility, the reallocation of employment towards industries with larger subsidies is stronger in

matter for instrument power; lagging shares beyond what is necessary would typically make the instrument weaker.

What constitutes the beginning of the natural experiment? If there were no shifts correlated with g_k in the past, it is just the beginning of the period when the g_k are measured. However, if the shifts unfold over several periods in a serially correlated way, it is appropriate to lag the shares further, to the first of these periods—or alternatively to extract unpredictable shock innovations and use them to construct the shift-share instrument. Another problem that arises with serially correlated shifts is that past shifts may have direct dynamic effects on the current outcomes (Jaeger, Ruist, and Stuhler 2018). Simply lagging the shares does not help with this problem. We discuss the problems arising in panels and possible solutions in online Appendix A.1.

5. Report descriptive statistics for shifts in addition to observations. Empirical papers normally present the number of observations and summary statistics for the main variables. In shift-share analyses that leverage the variation in shifts, it is important to also present such descriptive statistics for these—in the same way as one would in a non-shift-share setting at the shock level. While the mean and standard deviation of z_i is useful to know, so are the mean and standard deviation of g_k.

One detail here is that, as we show below, each shift carries an importance weight proportional to the exposure share of that shift for an average observation, $s_k = \frac{1}{N}\sum_i s_{ik}$. For example, when studying subsidy shifts across industries, the importance weights could correspond to the average industry employment share across local labor markets. Thus, it is natural to report descriptive statistics with those weights as well. For instance, the weighted version of the number of shifts is the "effective number of shifts"—the inverse of the Herfindahl index of shock importance weights, $1/\sum_k s_k^2$. When the effective number of shifts is small, a few shifts may drive the empirical analysis, potentially making the results noisy and unreliable. This is not specific to shift-shares: a similar issue can arise when running a weighted ordinary least squares regression, if some observations get disproportionately large weights.[6]

Descriptive analyses for the shifts need not be limited to their effective number and the distribution. For instance, one could also describe the distribution of the shifts after residualizing them on shift-level controls the researcher plans to include. Or one could plot the shifts on the map if they have a geographic dimension.

6. Implement balance tests for shifts in addition to the instrument. In every research design, it is useful to perform balance tests: specifically, to check that the variation

flexible local labor markets. If subsidies are random but persistent across the two periods, industries with large subsidies will be increasingly concentrated in regions with flexible labor markets. The shift-share instrument will therefore take higher values in flexible labor markets, causing bias if flexible labor markets also have stronger employment growth for other reasons.

[6] If shifts are correlated within certain clusters, the Herfindahl index can be computed at the level of such shift clusters, as having many correlated shifts may also not be enough for a reliable statistical analysis.

believed to be exogenous is indeed not correlated with proxies for confounders. In a shift-share design with exogenous shifts, this can be done in two ways: for the instrument at the level of units, and also directly for the g_k at the level of shifts.

Checking balance of the instrument at the unit level is relatively standard. For instance, a typical pre-trend test involves regressing the lagged outcome on z_i while including the controls picked in advance (such as the incomplete share control). The only particularity of shift-share designs in this case is that standard errors should be computed appropriately, as we discuss in the next step.

But when the identifying variation is at the shift level, it is also useful to check balance of shifts directly, with respect to shift-level variables that may proxy for unobservables. For example, in our running example with a change in industry subsidies, one could check whether the shifts correlate with variables reflecting labor supply factors, such as the composition of the workforce and the share of immigrants in the industry. This test is useful to assess whether changes in subsidies are systematically different for certain industries that would likely have been on different employment trends even absent changes in labor demand.

7. Produce the main estimates with correct standard errors and check sensitivity. Valid statistical inference in shift-share designs with exogenous shifts requires a special "exposure-robust" approach. Intuitively, inference must take into account that units with similar shares mechanically have correlated z_i and may also have correlated ε_i due to their common exposure to unobserved shocks. For example, regions specializing in the same industries will be affected by the same (potentially unobserved) industry shocks. Adão, Kolesár, and Morales (2019) show with a Monte Carlo simulation that this issue can be very serious in practice.

Two solutions have been developed, both leveraging as-if random assignment of the shifts. First, Adão, Kolesár, and Morales (2019) provide a variance estimator that is asymptotically valid regardless of the correlation structure of the errors across observations, as long as the exogenous shifts are mutually uncorrelated or clustered in a known way (for example, by group of industries). Second, Borusyak, Hull, and Jaravel (2022) show that one can simply run a particular shift-level two-stage least squares regression, which always produces an identical coefficient as $\hat{\beta}$ from the shift-share regression (1) but gives valid standard errors, because it is estimated at the same level at which the shifts are assigned. In this regression, the k-level outcome and treatment are certain transformations of the original outcome and treatment, shifts g_k directly serve as a single instrument, shift-level controls q_k are directly included as controls, and estimation is weighted by average shares $s_k = \frac{1}{N}\sum_i s_{ik}$.[7] The *ssaggregate* packages in Stata and R automate the transformation of the outcome and treatment for this regression. The shift-level regression offers the flexibility to accommodate various types of dependence in the shifts; for example, not only standard clustering but also spatial clustering and serial

[7] Specifically, the transformation of the outcome and treatment involves first residualizing them on the included i-level controls and then, for each k, averaging across observations with weights s_{ik}.

correlation. The equivalent regression can also be used to produce exposure-robust first-stage F-statistics to judge the instrument strength.

After producing the main shift-share estimates, it can be instructive to check their robustness to a variety of choices. For example, one may examine the stability of the estimate under alternative sets of controls which could correspond to different assumptions of conditional quasi-random shift assignment. Similarly, one may check that estimation with and without unit-level importance weights (for example, population weights in a regional analysis) yields similar results.

Examples of the Shift-Based Approach

We now discuss two examples, which illustrate some of the key practical insights for shift-shares with exogenous shifts. The first example focuses on how to use the shift-share design with a true experiment. The second describes a shift-share design with quasi-experimental shifts and illustrates why "incomplete shares" deserve special attention. Online Appendix A.1 provides an additional example leveraging time-series variation in the shifts.

Shift-Share in a Randomized Trial. Franklin et al. (2024) leverage randomized shifts in a shift-share design to estimate the indirect impacts of an intervention. They study a large public works program offering employment at high wages to low-income workers residing in specific neighborhoods in Addis Ababa, Ethiopia. The authors estimate the impact of this program on private sector wages: by increasing employment in public works, the program can reduce labor supply for other activities and increase private wages. Identification relies on the randomized rollout of the program, and the authors find large wage effects.

While the program is randomized at the level of residential neighborhoods k, it may have spillovers on wages in other neighborhoods (labor markets i) because workers can commute. Using data on baseline-period commuting, Franklin et al. (2024) build a measure of each labor market's exposure to the randomized roll-out: for each labor market, the shift-share treatment takes an average of intervention dummies across places of residence (the shifts) weighted by the share of workers who commute from those places of residence (exposure shares which sum to one).

In this setting, if the shifts are simply randomly assigned, there is no need to introduce controls. Imagine, however, that some residential neighborhoods k were ineligible for randomization. Then, the total share of commuters from eligible areas is less than one, and controlling for this total is necessary. With this control, and assuming that commuting shares correctly capture the structure of spillovers, the shift-share design identifies the causal impact of the program.

Shift-Share without an Experiment. Autor, Dorn, and Hanson (2013) study the impact of import competition with China on US employment. While this relationship could be analyzed across industries, they adopt the "local labor markets approach" (following, for example, Topalova 2010). Simplifying details, they define the outcome as the employment change in a US local labor market (commuting zone) and the treatment as the change in local exposure to import competition. Local exposure is measured as the average of national industry changes of imports from China (in dollars per US worker), weighted by local employment weights of different industries.[8] Ordinary least squares estimates may be biased if, for example, high productivity growth in China happens in industries with systematically different productivity or demand trends in the United States or if US consumers substitute to Chinese goods in industries where US productivity is lagging.

In this setting, the idealized experiment would be to assign observed productivity shifts at random across manufacturing industries in China. These shifts would have different incidence across US commuting zones given the predetermined industrial composition of each area. In practice, productivity changes are unobserved and, as a proxy, Autor, Dorn, and Hanson (2013) use the observed growth of imports from China in industry k in eight high-income countries (excluding the United States). Measuring imports in those countries ensures that demand and supply shocks that are idiosyncratic to the US cannot bias the results.

An important feature of this setting is that the exposure shares do not sum to one, because only manufacturing industries are exposed to trade with China. Locations with a larger total share of employment in manufacturing are likely on different potential outcome trends, for example, because of the secular decline in manufacturing (which can have many causes other than trade). To address this issue, it is necessary to control for the sum of exposure shares in each location. Note that the appropriate control equals the total regional share of manufacturing employment in the period in which the shares are measured. Because Autor, Dorn, and Hanson (2013) lag the shares by a decade relative to the period of the outcome and treatment, the incomplete share control should be lagged as well.

A further adjustment is called for because Autor, Dorn, and Hanson (2013) conduct the analysis in a repeated cross-section over two ten-year periods, and the average shifts are different in the two periods. Here, leveraging shock variation across industries only within periods requires controlling for the interaction of the sum of exposure shares with period indicators. This control prevents the bias that would arise if the manufacturing sector as a whole (and regions specializing in manufacturing industries) declined at different rates in the two periods for reasons unrelated to trade.

To assess the plausibility of the design, it is instructive to conduct industry- and commuting-zone-level balance tests. At the industry level, it could be that China specializes in certain industries (for example, low-skill industries) that could have been on different employment trends in the US absent trade shocks. To speak to this concern, one can correlate the shift g_k with industry-level variables reflecting the structure of employment and technologies—such as the skill and labor intensity,

[8] This approach is meant to account for important spillovers across industries: if workers can move from an industry affected by import competition to another one, declines in industry employment are not informative of the aggregate effects of import competitions. Spillovers across commuting zones are likely more limited.

average wages, and investment in new technologies (for example, computers) in a pre-period. Using the data from Autor, Dorn, and Hanson (2013), Borusyak, Hull, and Jaravel (2022) find that the shifts are balanced across these dimensions.

Correlating the regional shift-share instrument with potential commuting -zone-level confounders is also instructive. Such a correlation would arise if China specializes in industries that are located in commuting zones with unusual observed characteristics, which can raise concerns they are on different potential employment trends, too. One can regress commuting-zone-level predetermined variables—such as the lagged fraction of population who is college-educated, foreign-born, female, or working in routine occupations—on the shift-share instrument, controlling for the sum of shares interacted with period fixed effects. One can also implement a standard "pre-trend" test, with lagged commuting zone outcome on the left-hand side. Autor, Dorn, and Hanson (2013) and Borusyak, Hull, and Jaravel (2022) find that most of these tests pass.

Using this shift-share strategy, Autor, Dorn, and Hanson (2013) document a substantial decrease in both manufacturing employment and total employment in local labor markets that were more exposed to import competition from China. Introducing incomplete share controls interacted with period indicators, Borusyak, Hull, and Jaravel (2022) find smaller effects, especially for total employment.

Exogenous Shares in Practice

We now provide a list of practical steps to determine whether and how to use the exogenous-share approach to shift-share designs. As before, these steps can also serve as a blueprint for readers of papers using these designs. A summary of the key takeaways is given in column 2 of Table 2. We develop this checklist with the immigration setting from above, where s_{ik} represents the lagged share of immigrants from country k in region i. We discuss several applied examples at the end of this section.

A Checklist for the Share-Based Approach

1. Determine whether the exposure shares are potentially suitable instruments. Like before, the researcher can start by motivating the outcome equation and describing the main sources of treatment endogeneity. With the exogenous shares approach in mind, the researcher would then provide reasons why the shares may be useful instruments to address the corresponding threats. While we illustrate how this can done with detailed empirical examples below, here we highlight two general guiding principles.

First, the instrument exogeneity argument requires shares to be "tailored" to the treatment. Recall that the shares cannot be exogenous instruments if they capture the exposure of the outcome to some unobserved shocks. This rules out cases where the shares are "generic," in the sense of capturing exposure to many shocks, while the treatment only captures one such mechanism (for example, import competition in Autor, Dorn, and Hanson 2013) and it is not feasible to control for the effects of all other shocks. Conversely, in our running migration example, it is conceivable that the share of migrants from a certain origin only captures the region's exposure to migration shocks, making such shares potentially exogenous.

Second, the identification strategy can be strengthened by exploiting a source of variation in the initial shares that is more likely to satisfy exogeneity. Terry et al. (2023), for instance, study the effects of migration on innovation and worry that the initial composition of migrants may be correlated with labor demand factors. For instance, migrants from certain origins may have settled in regions with persistently strong labor demand, which could directly impact innovation. They address this issue by replacing the shares with their component arising from specific historical quasi-experiments, leveraging how the timing of historic waves of immigration coincided with the timing of growth across US regions.

A simpler strategy of lagging the shares can also sometimes help, but it does not by itself guarantee exogeneity. Lagging the shares will typically weaken the instrument, so it is important to explain why it is plausible that lagging the shares reduces their covariance with the error term by more than it reduces the covariance with the treatment. For example, in studies of the effect of immigration on local labor markets, lagging the shares to an earlier decade is helpful if (1) labor demand shocks that attract migrants are transitory, and (2) new migrants persistently go to places where migrants from the same origin arrived earlier.

2. Choose the necessary unit-level controls. Even if the shares are tailored, their exogeneity is a nontrivial assumption—as with any parallel trends assumption. As usual, exogeneity can be relaxed by including control variables. For example, the researcher can control for certain sums of shares to only leverage share variation conditional on these sums. In the migration setting, controlling for the initial total immigrant share would mean that the shift-share would leverage variation in the composition of migrants across locations, avoiding comparisons between regions with high and low migration intensity overall.

3. Characterize which shares matter the most for the estimates. When viewing the shift-share estimate as a pooled version of K one-at-a-time share-instrument estimates, it can be important to understand whether a small subset of these instruments drive the results. If this is the case, the researcher can use those shares to explain how the identification strategy works and can focus on them for the balance tests described below.

Goldsmith-Pinkham, Sorkin, and Swift (2020) show how to measure the importance weight of each share instrument, which they refer to as "Rotemberg weights" (referencing Rotemberg 1983). They are based on a decomposition of the shift-share estimator into a weighted sum of individual-share-instrument estimators with weights that add up to one, although some can be negative. These weights are larger for shares that are exposed to a bigger g_k and that are more predictive of treatment. The Rotemberg weights can be interpreted as measuring the sensitivity of the shift-share estimate to violations of exogeneity by each share instrument. The *bartik_weight* command in Stata and R provided by Goldsmith-Pinkham, Sorkin, and Swift (2020) computes these weights.[9]

4. Implement balance tests for individual shares in addition to the instrument. Like in any design, it is worth checking balance of the instrument on observable variables that may be expected to correlate with the error term. Different variables can serve for useful balance tests: pre-period changes in the outcome variable (corresponding to a pre-trends test), unit characteristics measured at the beginning of the period, or contemporaneous changes in placebo outcomes that are not expected to be causally affected by the treatment.

The special feature of the exogenous-share approach is that balance tests can also be performed on individual shares, as each of them is assumed to be exogenous. To avoid issues with testing many hypotheses, it is natural to focus on the subset of shares that are most important for the resulting estimate as measured by the Rotemberg weights. We note that outcome pre-trends are more likely uncorrelated with individual shares when either the shares have changed drastically since the pre-period or there were no shocks of the same nature as g_k in the past (Jaeger, Ruist, and Stuhler 2018).

5. Check sensitivity to how share instruments are combined. When shares are exogenous, the parameter β is *overidentified*: any individual share or linear combination of shares is itself a valid instrument. The shift-share instrument is one such combination but, because many others are available, it is instructive to check that the shift-share estimate would not be too dependent on the researcher's choice. Here we review several such tests and discuss what their failure may indicate.

A standard statistical test for whether using each of the individual shares as an instrument yields statistically indistinguishable estimates of β is the Sargan-Hansen overidentification test (Wooldridge 2002, Ch. 6.2.2). Graphical procedures aid the interpretation of this test. A conventional "visual instrument variable" procedure (Angrist and Pischke 2008, p. 103) plots K reduced-form coefficients (from regressions of the outcome on a given share, including all controls) against corresponding first-stage coefficients (from similar regressions of the treatment). Because individual-share estimates of β are given by the ratio of reduced-form and first-stage coefficients, the points in this plot should all lie on a single ray from the origin when all share instruments estimate the same parameter (that is, when the overidentification test "passes").[10] An alternative graph is proposed by Goldsmith-Pinkham,

[9] Unfortunately, the Rotemberg weights are not unique when the shares add up to one. This is because the shares—and thus individual-share estimators—are perfectly multicollinear. In this case, Goldsmith-Pinkham, Sorkin, and Swift (2020) recommend choosing the Rotemberg weights that correspond to the demeaned shifts.

[10] Formally, online Appendix B.2 shows that the shift-share estimate of β equals the coefficient from a regression of reduced-form coefficients on first-stage coefficients, with no intercept and with particular weights related to the Rotemberg weights.

Sorkin, and Swift (2020): a scatter plot of the K estimates of β, each using one of the shares as an instrument, against their respective first-stage F-statistics. Here one hopes to see that all estimates are similar, especially those with high F-statistics and large Rotemberg weights.

Because individual share instruments may not be very strong, it is also useful to check the sensitivity of the β estimates to alternative combinations of multiple share instruments. One may examine whether the estimate changes when using only a few shares—for example, those with the largest Rotemberg weights. Another approach is to keep all shares for higher precision but combine them in a different way. When K is small relative to the sample size, two-stage least squares is the natural estimator to report; an efficient generalized method of moments estimator is another option. With many shares, two-stage least squares suffers from bias but several estimators robust to "many weak instruments" are available instead: jackknife instrumental variables (Angrist, Imbens, and Krueger 1999), limited information maximum likelihood, the heteroskedasticity-robust Fuller estimator (Hausman et al. 2012), and modified bias-corrected two-stage least squares (Kolesár et al. 2015).[11]

It is comforting if all of the above checks indicate robustness of the shift-share estimate; but what should the researcher conclude if not? The answer depends on whether the causal effect of x_i on y_i can vary across units i (making the constant-β model (1) misspecified). When the effects are homogeneous, the failure of the above tests indicates that the share exogeneity assumption is violated. This need not be the case with heterogeneous effects, as different combinations of share instruments may estimate different combinations of causal effects even when all share instruments are exogenous. Still, sensitivity of the estimates to the choice of the instruments is a cause for concern: the interpretation of shift-share estimates (and those from the alternative estimators) can be challenging under such effect heterogeneity. They may not represent the average effect for some subpopulation of "compliers," especially because share instruments are correlated with each other (Mogstad, Torgovitsky, and Walters 2021).

Examples of the Share-Based Approach

We now use two examples to illustrate the exogenous-share approach in the contexts of the labor market responses to migration and retirement rates. Both examples show the tight conceptual link between the exogenous shares approach and difference-in-differences research designs.

Labor Market Effects of Immigration. We first consider the design of Card (2009, Table 6) and its reanalysis by Goldsmith-Pinkham, Sorkin, and Swift (2020). The goal is to estimate the (inverse) elasticity of substitution between immigrant workers and native workers in labor demand, that is, the relationship between the

[11] The shift-share estimator also requires a similar bias correction when the shifts are estimated from the sample, as in Bartik (1991). In online Appendix A.12, we discuss the leave-out shift-share estimator that helps in this scenario.

log wage gap between immigrant and native workers and the ratio of immigrant to native hours worked. Simplifying details, the analysis considers a cross section of outcomes (in levels) in 2000 across 124 cities, separately for high-school and college-educated workers.

As with any demand equation, ordinary least squares estimates may be biased: a positive labor demand shock for migrants would draw more immigrants into a location and at the same time increase their wages relative to natives. An instrument is needed that shifts the relative supply of migrant and native workers. Card (2009) proposes a shift-share instrument, leveraging immigration patterns from 38 countries indexed by k. Here s_{ik} is the share of immigration group k in the population of city i in 1980; note that these shares add up to the initial migration share rather than one (and the initial migration rate is not controlled for). The g_k is the number of migrants in group k moving to the United States from 1990 to 2000, normalized by the national stock of migrants from k already in the United States in 1990.

This shift-share strategy alleviates some endogeneity concerns, as the shares are uncorrelated with some relative labor demand factors. Specifically, transitory regional labor demand shocks (which attract migrants to a particular location in the current period only) would be a problem for ordinary least squares but not for Card's instrument, as the migrant shares are measured before these shocks are realized. In contrast, persistent regional labor demand factors (for example, characteristics that always make immigrants more productive relative to native workers, such as the prevalence of certain languages like Spanish) would remain a problem for both ordinary least squares and the shift-share approach, as these factors impact the beginning-of-period migrant share while also entering contemporaneous labor demand in the error term.

Some of the potential limitations of Card's instrument can be addressed by simple adjustments to the empirical strategy. In particular, estimating the outcome equation in differences would alleviate concerns about time-invariant regional confounders. Moreover, controlling for the total initial share of migrants would make the shift-share leverage the *composition* of migration origins. This would address labor demand shocks that affect all migrants equally.

Working with the original Card (2009) setting, Goldsmith-Pinkham, Sorkin, and Swift (2020) compute the Rotemberg weights to show which shocks matter most for the estimates. They show that Mexico receives half of the weight in the sample of high school equivalent workers. Thus, for these workers one can largely think of the research design as using the initial Mexican immigrant share as the instrument. Indeed, Card (2009) notes that the shift-share is highly correlated with the initial fraction of Mexican migrants. For college equivalent workers, Goldsmith-Pinkham, Sorkin, and Swift (2020) document that the top country is the Philippines, receiving 15 percent of the total weight.

Goldsmith-Pinkham, Sorkin, and Swift (2020) also perform balance tests for share instruments with high Rotemberg weights. They report that the 1980 Mexican immigrant share does not predict relative wages in 1980 or 1990, but does in 2000 (the year of analysis). While the patterns for Mexico are comforting, the 1980 share of immigrants from the Philippines correlates with the native-immigrant wage gap in all three periods. Other countries also feature statistically significant violations of pre-trend balance, raising concerns about share exogeneity. In part, the correlation between pre-period outcomes and certain origin shares could arise because pre-period outcomes are affected by pre-period immigration rates. Including lagged immigration rates in the model could help pre-trend tests pass while also making causal estimates more credible.

With these caveats, shift-share estimates suggest that when the ratio of immigrant to native hours worked increases by 10 percent (because of supply shocks), the wage of migrants relative to native workers falls by 4 percent for high school graduates, and by 7 percent for college graduates. This implies that migrant and native workers are more substitutable for low-skill groups. Goldsmith-Pinkham, Sorkin, and Swift (2020) also report the results obtained with alternative estimators, such as two-stage least squares. They find that the point estimates remain very similar. Similarly, plotting the estimates using individual shares as instruments against their respective *F*-statistics, they find little variation in the estimates, especially for the strong instruments. In online Appendix Figure 1, we report the visual instrumental variables graph. Because individual estimates are similar for all origin countries, the estimates lie near the ray through the origin with the slope equal to the shift-share estimate. These tests demonstrate the robustness of the baseline estimate to alternative ways of combining the share instruments and indicate that treatment effect heterogeneity is a limited concern in this setting.

Labor market effects of retirement. Mohnen (2025) studies the impact of the retirement rate of older generations on labor market outcomes for younger generations in the United States, conducting the analysis at the commuting zone level. Specifically, the author relates ten-year differences in labor market outcomes for the young (unemployment rate, share working in high-skilled jobs, and so forth) to retirement rates over ten years in the commuting zone. The specification further includes start-of-period regional controls: employment share of manufacturing and routine occupations, unemployment rate, and so forth. Still, ordinary least squares estimates may be biased because strong labor demand in some commuting zones may explain both low retirement rates among older workers and low unemployment rates for younger workers.

The author addresses this identification challenge with a shift-share strategy that leverages cross-area variation in age composition among the older population. Specifically, the instrument for the ten-year retirement rate in commuting zone i uses the local share of age k among the population aged 45 to 80 as s_{ik} (such that shares sum to one in each commuting zone), and the national ten-year retirement rate by age as g_k. The age composition predicts retirement rates because older workers are more likely to retire, giving the shift-share instrument power. The identification assumption is that the age shares (among people above 45) are all valid instruments conditional on the controls.

To better understand the source of variation, the author describes which age shares matter the most in driving the estimates. He reports Rotemberg weights, documenting that they are close to proportional to the ten-year national retirement rates by age.

Shift-share estimates suggest that the retirement slowdown in the United States in recent decades was detrimental to career outcomes for the youth. In places where fewer workers retire, young workers have lower wages and are more likely to have low-skill jobs, and their job mobility falls, although their unemployment does not increase.

To assess whether the results might depend on how the share instruments are combined, the author performs an overidentification test, which passes. He also reports alternative estimators: using a particular combination of shares (the initial share of the population age 52–59 as a fraction of the population above 45), or all detailed age shares as separate instruments via generalized method of moments. All estimates suggest similar results, lending support to the validity and robustness of the design.

Conclusion

We have reviewed two frameworks for shift-share research designs, which include sufficient conditions for instrument validity, narratives for interpreting these conditions intuitively, balance tests for the assumptions, and various practical recommendations. Table 2 summarizes the key practical insights for the two approaches, leveraging either exogenous shifts or shares.

How can one pick between the two approaches? In some settings one approach is a "non-starter"; for example, the exogenous shifts approach with too few shifts or the exogenous shares approach when the treatment is specific while the shares are generic. In other settings, it may be productive to think through the potential bias and efficiency properties of the instruments each approach would suggest. For instance, when estimating the local demand elasticity for migrant labor, can a plausibly exogenous supply shift ("push factor") with a strong effect on migration be found? Or is it plausible that there are no national demand shifts for migrants of any origins—in which case a (likely stronger) share-based instrument may be convincing enough? We hope our review will help researchers assess such tradeoffs.

■ *We thank Matilde Bombardini, David Dorn, Michael Gmeiner, Paul Goldsmith-Pinkham, David McKenzie, Paul Mohnen, Jonathan Parker, Nina Pavcnik, Timothy Taylor, and Heidi Williams for helpful suggestions.*

References

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128 (6): 2188–2244.

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2023. "General Equilibrium Effects in Space: Theory and Measurement." NBER Working Paper 25544.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics* 134 (4): 1949–2010.

Aghion, Philippe, Céline Antonin, Simon Bunel, and Xavier Jaravel. 2022. "Modern Manufacturing Capital, Labor Demand, and Product Market Dynamics: Evidence from France." Programme on Innovation and Diffusion Working Paper 044.

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14 (1): 57–67.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Impacts of Import Competition in the United States." *American Economic Review* 103 (6): 2121–68.

Autor, David H., and Mark G. Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118 (1): 157–205.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* W. E. Upjohn Institute for Employment Research.

Berman, Nicolas, Antoine Berthou, and Jerome Hericourt. 2015. "Export Dynamics and Sales at Home." *Journal of International Economics* 96 (2): 298–310.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* 23 (1): 1–75.

Bombardini, Matilde, and Bingjing Li. 2020. "Trade, Pollution and Mortality in China." *Journal of International Economics* 125: 103321.

Borusyak, Kirill, and Peter Hull. 2021. "Non-Random Exposure to Exogenous Shocks: Theory and Applications." NBER Working Paper 27845.

Borusyak, Kirill, and Peter Hull. 2023. "Non-Random Exposure to Exogenous Shocks." *Econometrica* 91 (6): 2155–85.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies* 89 (1): 181–213.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2024. "Design-Based Identification with Formula Instruments: A Review." *Econometrics Journal*. https://doi.org/10.1093/ectj/utae003.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies* 91 (6): 3253–85.

Borusyak, Kirill, and Matan Kolerman-Shemer. 2024. "Regression Discontinuity Aggregation, with an Application to the Union Effects on Inequality." Unpublished.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet. 2015. "Social Networks and the Decision to Insure." *American Economic Journal: Applied Economics* 7 (2): 81–108.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *ILR Review* 43 (2): 245–57.

Card, David. 2009. "Immigration and Inequality." *American Economic Review* 99 (2), 1–21.

Christian, Paul, and Christopher B. Barrett. 2024. "Spurious Regressions and Panel IV Estimation: Revisiting the Causes of Conflict." *Economic Journal* 134 (659): 1069–99.

Clots-Figueras, Irma. 2011. "Women in Politics: Evidence from the Indian States." *Journal of Public Economics* 95 (7–8): 664–90.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum. 2014. "The Rise of the East and the Far East: German Labor Markets and Trade Integration." *Journal of the European Economic Association* 12 (6): 1643–75.

de Chaisemartin, Clément, and Xavier D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964–96.

de Chaisemartin, Clément, and Ziteng Lei. 2023. "More Robust Estimators for Instrumental-Variable Panel Designs, with an Application to the Effect of Imports from China on US Employment." http://dx.doi.org/10.2139/ssrn.3802200.

Franklin, Simon, Clément Imbert, Girum Abebe, and Carolina Mejia-Mantilla. 2024. "Urban Public Works in Spatial Equilibrium: Experimental Evidence from Ethiopia." *American Economic Review* 114 (5): 1382–1414.

Freeman, Richard B. 1975. "Overinvestment in College Training?" *Journal of Human Resources* 10 (3): 287–311.

Freeman, Richard B. 1980. "An Empirical Analysis of the Fixed Coefficient 'Manpower Requirements' Model, 1960–1970." *Journal of Human Resources* 15 (2): 176–99.

Goldsmith-Pinkham, Paul. 2024. "Tracking the Credibility Revolution across Fields." https://doi.org/10.48550/arXiv.2405.20604.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110 (8): 2586–2624.

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Nguyen. 2020. "Do Credit Market Shocks Affect the Real Economy? Quasi-experimental Evidence from the Great Recession and 'Normal' Economic Times." *American Economic Journal: Economic Policy* 12 (1): 200–225.

Hausman, Jerry A., Whitney K. Newey, Tiemen Woutersen, John C. Chao, and Norman R. Swanson. 2012. "Instrumental Variable Estimation with Heteroskedasticity and Many Instruments: Instrumental Variable Estimation." *Quantitative Economics* 3 (2): 211–55.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang. 2014. "The Wage Effects of Offshoring: Evidence From Danish Matched Worker-Firm Data." *American Economic Review* 104 (6): 1597–1629.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Jaeger, David A., Joakim Ruist, and Jan Stuhler. 2018. "Shift-Share Instruments and the Impact of Immigration." NBER Working Paper 24285.

Jaravel, Xavier. 2019. "The Unequal Gains from Product Innovations: Evidences from the US Retail Sector." *Quarterly Journal of Economics* 134 (2): 715–83.

Kolesár, Michal, Raj Chetty, John Friedman, Edward Glaeser, and Guido W. Imbens. 2015. "Identification and Inference with Many Invalid Instruments." *Journal of Business and Economic Statistics* 33 (4): 474–84.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1), 159–217.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R. Walters. 2021. "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables." *American Economic Review* 111 (11): 3663–98.

Mohnen, Paul. 2025. "The Impact of the Retirement Slowdown on the US Youth Labor Market." *Journal of Labor Economics* 43 (1). https://doi.org/10.1086/725874.

Nunn, Nathan, and Nancy Qian. 2014. "US Food Aid and Civil Conflict." *American Economic Review* 104 (6): 1630–66.

Rotemberg, Julio J. 1983. "Instrument Variable Estimation of Misspecified Models." MIT Sloan Working Paper 1508-83.

Terry, Stephen J., Thomas Chaney, Konrad B. Burchardi, Lisa Tarquinio, and Tarek A. Hassan. 2023. "Immigration, Innovation, and Growth." https://sciencespo.hal.science/hal-03869993v1.

Topalova, Petia. 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." *American Economic Journal: Applied Economics* 2 (4): 1–41.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Xu, Chenzi. 2022. "Reshaping Global Trade: The Immediate and Long-Run Effects of Bank Failures." *Quarterly Journal of Economics* 137 (4): 2107–61.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 205–224

Tax Privacy

Joel Slemrod

In Norway (and some other countries), any citizen can log on to a government website and learn the declared taxable income and income tax liability of any other Norwegian. This public policy seems unimaginable to most Americans. Even if taxpayer information stayed securely within the US Internal Revenue Service, some voices decry the extent of private information the tax agency now demands. Moreover, opposition on grounds of invasion of privacy is at the forefront of resistance to current policy initiatives such as an expansion of IRS enforcement activity, or the possibility of a personal wealth tax.

An equitable and efficient tax system requires that the government have access to certain information about taxpayers. If the tax liability and ultimately the tax burden of individuals and families is to depend on their level of well-being, then indicators of that level must be available. If liability is to depend on characteristics of taxpayers of given well-being, then indicators of those characteristics must be available to the government. To the extent that taxation of businesses is part of the tax system, information about their activities, and often their profits, is needed. Conversely, if the demand for privacy implies limiting government's access to information, it constrains the extent to which a tax system can be equitable and efficient. In this way, demand for limiting government access to information imposes social costs.

How much social cost is justified depends on the value placed on this privacy by people and firms. In many non-tax situations, individuals and firms voluntarily reveal information about themselves that is relevant to their tax liability—when it is in their interest to do so. For example, people often reveal their (reported taxable)

■ *Joel Slemrod is the David Bradford Distinguished University Professor of Economics, University of Michigan, Ann Arbor, Michigan. His email address is jslemrod@umich.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20241431.

income in applying for a mortgage, applying for financial aid for college, and so on. Since the advent of the Internet and smartphones, information revelation in general has exploded, with most people affected not knowing the details of what is revealed and how that information is used. People may reveal their location via the GPS tracking on their phone, their interests via the terms they search for, their consumption habits via their online purchases, and their social connections via who they "friend."

Most Americans are at least vaguely aware of this information transfer, and a majority assert they believe the gains from these activities do not exceed the costs. For example, in a 2019 Pew Research Center poll (Auxier et al. 2019), 81 percent of Americans said that the potential risks they face because of data collection by companies outweigh the benefits, while 66 percent said the same regarding government data. However, many people do relatively little to protect their privacy; for example, they post personal information on social media and do not update computer passwords. This pattern of responses has given rise to the idea of a "privacy paradox," whereby professed worry about privacy does not translate into action.[1]

Tradeoffs concerning the privacy of personal information have figured prominently in public debates over tax policy, and have affected both the acceptability and the design of taxes. The second, and more successful, version of the modern income tax, introduced in the United Kingdom in 1803 by Prime Minister Henry Addington, differed from its less successful predecessor in that it was "schedular"; that is, tax was applied not to the sum of income of all types, but separately to five different categories of income (as discussed in Keen and Slemrod 2021, pp. 111–12). The purpose of the schedular tax was not to apply different tax rates to the different types of income, but rather to preserve an element of privacy: it meant that no revenue officer could know a taxpayer's total income. A century earlier, England replaced a tax on hearths with a tax on windows, in part because the latter did not require the tax collector to enter taxpayers' homes.[2] In modern times, public arguments over issues like a wealth tax or greater enforcement resources for tax authorities are often phrased in concerns over reduced privacy. This essay reviews the issues and the state of knowledge on tax policy and privacy concerns.

What Is Tax Privacy? External and Internal Aspects

In practical terms, tax privacy refers to two different, but related, categories: (1) the ability of the government to prevent the information it accesses during the tax process from leaking beyond the tax authority, and (2) the ability of people not to have to reveal certain information to the government as part of the tax process. We might refer to these as external and internal aspects of tax privacy.

[1] The existence of this "paradox" has been challenged by, among others, Acquisti, Brandimarte, and Loewenstein (2015).

[2] Adam Smith, in *The Wealth of Nations* (1776, Book V, Ch. II, Pt. II, Art. I), describes it thus: "This odious visit rendered the [hearth] tax odious. Soon after the revolution, it was abolished as a badge of slavery."

There are rules in place to constrain external privacy. The IRS Taxpayer Bill of Rights lists both a right to privacy and a right to confidentiality.[3] In addition, IRS Code Section 6103 prohibits the release of any tax information by an IRS employee, subject to certain exceptions, which include sharing information with state tax agencies; sharing information with law enforcement agencies for investigation of non-tax criminal laws; official tax administration investigations; and sharing information with the Social Security Administration to establish tax liability. A criminal statute (§7216) prohibits preparers of tax returns from "knowingly or recklessly disclosing or using tax return information." This provision was enacted in 1971, and regulations updating it in light of the growth of electronic filing and the cross-marketing of products and services were effective beginning in 2009. Tax preparers can only use the information provided to complete and file one's tax return, unless the taxpayer stipulates otherwise.

One can imagine a scenario in which transmission of information to the tax agency triggers little or no concern over external tax privacy—if the information is transmitted digitally, safely stored, and accessed only to check its accuracy and to trigger communication with the taxpayer if further payment is due or further information is required. These steps can all be carried out by computer, as is the case with what the IRS calls "correspondence audits." If a "field audit"—a face-to-face examination—is deemed necessary, then the originally transmitted information and any further information requested will be known to the individuals on the auditing team. In either scenario, no—or few—individuals have access to taxpayer information. If those with access are bad actors, and violate the secrecy rules of IRS employees, then violation of tax privacy can occur.

However, this scenario may not apply, in a way that taxpayer information becomes much more widely disseminated. In what follows I discuss two examples, one an intended policy choice and the other a violation of law.

Public Disclosure

In several countries, at least some information from income tax returns of both individuals and companies is made available to the public as a policy choice. Public access to corporate tax information is provided in Australia, Iceland, Finland, Norway, Pakistan, and Sweden, and was once the law in Japan. Personal-level public disclosure is available on the internet in Norway, whereas Sweden, Finland and Iceland have systems where one can apply to the tax authorities for information about individuals, in Iceland for only a very limited time period. Pakistan introduced public disclosure for both corporate and personal tax information in 2012. Moreover, the issue is on the policy agenda in several countries, including almost half of the OECD countries. Nakayama (2021) provides a helpful overview of current public disclosure systems, including general disclosure policies such as discussed above, targeted disclosure programs that deal with making public the identities of delinquent taxpayers (as

[3] This document is accessible at https://www.irs.gov/taxpayer-bill-of-rights.

many US states have) or those that have been indicted or are being investigated. In a few countries, such as Japan and Turkey, programs publicly honor "good" taxpayers, and thereby reveal something about their tax remittances.

Even the United States has had public disclosure of income tax information, if only with occasional and brief experiments. The short-lived Civil War-era income tax provided that the public would be entitled to see the names and tax liabilities of taxpayers. The corporate excise tax of 1909 contained a publicity provision, which was soon after repealed. Information from tax year 1923 income tax returns was publicly released in 1924, although this lasted only to the provision's repeal in 1926. This episode gave rise to one of the most vivid privacy-related objections to public disclosure. Senator Louis Murphy (D-IA) stated at the time that disclosing income tax data is equivalent to taking "the curtains and shades from the homes of our taxpayers and pull[ing] out the walls of the bathroom to assure that the Peeping Toms shall have full and unobstructed opportunity to feast their eyes on the [tax return]" (as quoted in Leff 1984, pp. 70–71). The Revenue Act of 1934 required all income tax filers to submit a pink-colored form that contained information from the return—name, address, gross income, deductions, taxable income, and tax liability—that would become public. This requirement was, though, abolished before it went fully into effect. These US dalliances with public tax disclosure are discussed in Kornhauser (2018).

Corporate information disclosure is a central topic in accounting research, and tax disclosure—both to the public and to tax authorities—is a burgeoning topic within that area. Required corporate tax disclosures have dramatically increased in the past decade, due in part to the rapidly increasing ability of tax authorities to process and link vast amounts of data (as reviewed in Hoopes, Robinson, and Slemrod forthcoming).

In all of these cases, public disclosure of tax information is generally promoted as a measure to strengthen transparency, equity, and accountability, but is also decried as an invasion of privacy that engenders envy and promotes "salary snooping" by colleagues and neighbors (Purcell and Rossi 2019). Each November 1, the day the Finnish government publishes citizens' income and tax payments, is known there as "National Jealousy Day." Reck, Slemrod, and Vattø (2022) document who searches for whose tax information in Norway, and speculate on the motivations for these searches. Perez-Truglia (2020) argues, using survey data and Norwegian tax records, that the higher wage and salary transparency due to public tax disclosure increases the gap in happiness and life satisfaction between richer and poorer individuals. This finding is reminiscent of Varian's (2009, p. 8) argument that neighbors may care about the assessment of my house not because they care about my assessment per se, but because they care about *their* assessment.

Hacking of Tax Agency Files

Disclosure of tax information can also happen due to unauthorized data leaks from the tax authority. In 2015, a data breach at the Internal Revenue Service resulted in more than 700,000 Social Security numbers and other sensitive information being exposed. In 2007, a major data breach in the United Kingdom at HM Revenue & Customs (HMRC) exposed the personal details of over 25 million people, including names, addresses, and National Insurance numbers. In 2020, HMRC was targeted by a phishing attack that resulted in the theft of login details for over 100,000 accounts. In June 2021, ProPublica published confidential, private and legally-protected US taxpayer information, naming particular individuals in a series of articles. ProPublica claimed it had obtained a "vast trove of Internal Revenue Service data on the tax returns of thousands of the nation's wealthiest people, covering more than 15 years" (Eisinger, Ernsthausen, and Kiel 2021). On September 29, 2023, the US Justice Department charged a former IRS contractor, Charles Littlejohn, with disclosing this tax return information, a felony that carries a maximum sentence of five years. He pleaded guilty to disclosing tax return information without authorization, and on January 29, 2024, was sentenced to five years in prison.

How might the prospect of such leaks have affected taxpayer behavior? I know of no evidence that sheds light on this question. Some evidence does, though, suggest that concern for information leaks can affect private companies. After it was revealed that three million European users had abandoned Facebook after data leaks in 2018, its shares fell by 19 percent (Neate 2018). Below I discuss evidence of how data protection rules have affected data usage.

What Information Do Taxpayers Not Want Revealed, and to Whom Do They Not Want It Revealed?

The kind of information that people consider sensitive, and which they do not want others to be able to access, will naturally vary across individuals. For many, it includes details that can be used to identify an individual, such as names, addresses, phone numbers, Social Security numbers, and passport information. It also covers financial information: data related to financial accounts, credit card numbers, bank account details, credit ratings, and transaction histories, in part due to their potential use in identity theft. Details about one's employment, salary, job performance, and work-related activities may also be considered sensitive. Finally, marital and family status, household members, and sexual orientation might be considered private.[4]

Some kinds of sensitive information are—at least under the current US tax regime—unlikely to be requested by and transmitted to a tax agency. In this category I put location information, online activities (although maybe once needed for purposes of US states that imposed "use tax" on items used in the state but purchased outside the state), contents of communications, and social media data. Other information is not directly requested, but could be revealed/inferred from

[4] Marital status is not easily observable in many states. For example, in California it is possible to obtain a confidential marriage license. I am grateful to Joshua Blank for pointing this out to me.

tax return information. In this category, I put health information, including susceptibility to addiction (from medical deductions in tax records) as well as political and religious beliefs (from charitable donations).

Taxpayers may also vary in privacy concerns over who has access to their information. In some cases, the issue may be access by other branches of the government, such as immigration authorities. Some taxpayers might not want their relatives (or spouses or lovers) to know their income or wealth. The auditing and accounting firm EY tells its clients that sensitive tax information can be bought and sold in certain countries, posing risks such as extortion and kidnapping for wealthy individuals: "Unwanted exposure can lead to invasive media attention, sensationalism and potential reputational damage" and "[E]nsuring privacy is crucial for preserving personal and professional standing" (Parets 2023).

Harassment of those that public disclosure revealed to be affluent was apparently a major reason that Japan eliminated its public tax disclosure system in 2004 (Hasegawa et al. 2013), and concern about kidnapping and extortion risk was a factor in the demise of the 1930s US pink-slip provision discussed above (Kornhauser 2018). One can imagine certain taxpayers not wishing potential employers or banks to learn details of their income or assets.

Some taxpayers undoubtedly are concerned with how the government might use the information in ways that go beyond collecting of taxes. Much of the academic writing about tax privacy implicitly assumes that the government cares only about tax information for purposes of collecting taxes, but there have been several episodes in which tax audits were directed for political purposes. In 1971, President Richard Nixon vowed to appoint as IRS Commissioner someone who would "do what he's told, that every income tax return I want to see I see" (according to taped transcripts reported by Langer 2014). The 1974 articles of impeachment charged Nixon with abusing power in unlawfully using the Internal Revenue Service for purposes not authorized by law, although the new commissioner apparently did not target those on Nixon's "enemies list."

In 2013, it was reported that the IRS had selected political groups applying for tax-exempt status for intensive scrutiny based on their names or political themes. On one side, a report commissioned by the Obama administration and later released by the Treasury Inspector General (2017) found that, from 2004 to 2013, the IRS used both conservative and liberal keywords to choose targets for further scrutiny. On the other side, under the Trump administration in 2017, the US Department of Justice apologized for targeting conservative organizations and paid a financial settlement (US Department of Justice 2017).

In the polarized US political landscape, the threat perceived by some observers of a future authoritarian regime raises the specter of an intriguing political crosscurrent. Until now, it has been mainly voices of the left that have argued for the expanded information requirements of a more progressive tax system, such as would be needed for a wealth tax. But since the reelection of President Trump in 2024, the left may become more be wary of granting tax authorities access to detailed personal information. According to his former chief of staff, John F. Kelly, President

Trump during his first term wanted several of his perceived political enemies to be investigated by the Internal Revenue Service (Schmidt 2022).

From whom tax information is to be protected can affect the structure of a tax withholding system. Consider that, in the United Kingdom, "exact withholding"—that is, tax withholding where no refund or balance is due at year-end—is sustained by the employee providing the government the tax-relevant personal information, such as number of dependents. The government then conveys the withholding implications of those circumstances to the employer by issuing a code, but not telling the employer how the code is calculated, and thus not revealing which personal circumstances generate the adjustment in withholding. The existence of this system suggests that, at least in the United Kingdom, employees are more concerned that their employer not learn their personal circumstances than the government have access to such information. With whom the government shares information can also have incidence effects. Garriga and Tortarolo (2024) conclude that, when Argentina shifted the disbursement of child benefits from employers to the government, employers lost between 6 and 14 percent of the transfers they had captured under the former system through paying lower wages to workers with children. This outcome is an example of the failure of the "remittance invariance folk theorem"—that who remits tax (or, in this base, benefits) is irrelevant to incidence—as discussed in Slemrod (2008) and empirically investigated in Kopczuk et al. (2016).

Some Economics of Tax Privacy

A vast literature expounds on different concepts of privacy. Lessig (1999), writing about information generally and not about tax policy, distinguishes three separate aspects of privacy: the burden of intrusion, the offense to one's dignity, and the "substantive" invocation of privacy concerns as a way to constrain the power of the state to regulate. Stuntz (1995, p. 1026) illustrates the last aspect with an example far from taxation—the use of contraception: "Just as a law banning the use of contraceptives would tend to encourage bedroom searches, so also would a ban on bedroom searches tend to discourage laws prohibiting contraceptives." If the means of enforcement are limited, so too will be the effectiveness of laws that require such enforcement. From a libertarian viewpoint, Fulda (1996) argues: "The right to privacy is essential to the preservation of freedom for the simplest of reasons. If no one knows what I do, when I do it, and with whom I do it, no one can possibly interfere with it."

As noted at the start of the paper, the desire to restrict the government from having access to certain information is in direct conflict with achieving an equitable distribution of the tax burden. The prevailing—although certainly not only—principle for allocating tax burden across individuals is to do so on the basis of "ability to pay," and in practice this is usually proxied by a combination of income and family circumstances (like number of dependent children). In the seminal modern treatment of optimal income tax progressivity due to Mirrlees (1971), people differ in their endowment of ability, which determines their wage rate. But in this model, ability is not observable to the government, so tax liability cannot be based on it. Instead, tax liability can be based on income, which is equal to the wage rate multiplied by the person's choice of labor supply. Although income is generally correlated with ability, as a tax base it is inferior to ability in two ways. First, it depends on individual choice, and this choice will be affected by an income tax—those facing higher taxes have an incentive to supply less labor. Second, it assigns tax liability based in part on personal tastes about leisure versus the goods that money buys: "leisure lovers" will have lower income and therefore lower tax liability merely because of their tastes.

In the Mirrlees (1971) model, income is observable without cost to the government. But in this essay, I am suggesting that there may nevertheless be a social cost, related to privacy, of revealing income. However, restricting government's access to indicators of well-being hampers implementing optimal tax system progressivity. It may also hinder meeting goals like horizontal equity; for example, while the ideal tax system based on "ability to pay" would allow deductibility of involuntary health expenses, (many) people are loath to reveal them out of concern about protecting their own privacy.

One potential alternative would be to enact laws based on ability to pay through impersonal taxes that reveal nothing about a taxpayer that creates a record. For example, a set of proportional excise taxes on luxury goods (that is, those goods with high income elasticities) and subsidies for goods and services with low income elasticities would deliver a progressive distribution of tax burden. However, creating progressivity in this way would come at the cost of some horizontal inequity, as it would, for example, burden the middle-income family with expensive tastes more than an approximately progressive tax based only on income. Nor could an impersonal tax system achieve the granular progressivity of an income tax, as it is constrained by the details of the income elasticities of taxable commodities. For example, if all households in the top 1 percent spend the same percentage of their income on yachts, a yacht excise tax cannot differentiate the proportional liability within that group. Finally, purchases of some luxury goods may not be entirely private: in Greece and Italy, police have passed along ownership information of expensive cars and yachts to their tax authorities, which then use this information to ascertain whether the person has reported sufficient income to justify ownership of such expensive goods.

A burgeoning recent theoretical literature has examined various aspects of privacy. For example, Pourbabaee and Echenique (2024) consider public goods provision when privacy concerns may add to people's reluctance to reveal their true preferences, and address how a rule that adds random noise to people's reported valuations can preserve privacy while minimally compromising its optimality. Krähmer and Strausz (2023) study how a monopolist's optimal pricing policy depends on the presence of data-sensitive consumers who incur a utility loss whenever they make a purchase from which the monopolist deduces something about their valuation, and stress that the result—including how net gains accrue to data-sensitive versus nonsensitive consumers—depends crucially on whether consumers' privacy preferences are public information or not. Finally, Dekel et al. (2022) introduce the concept of a "privacy elasticity," which measures how behavior responds to a measurable change in privacy, and estimate this elasticity in a public-good game in a controlled lab environment, measuring how contributions change in response to marginal privacy conditions. This exercise may augur the rigorous marriage of optimal taxation and optimal privacy, facilitated by a sufficient statistic for privacy to accompany the central sufficient statistics of taxation such as the elasticity of taxable income. In sum, a recent literature on mechanism design and privacy has looked at problems related to optimal taxation in the presence of privacy concerns, but not at tax privacy directly. I would welcome efforts by public finance theorists to engage with it and address the policy tradeoff between guarding tax privacy and the achievement of other social objectives.

How Much Do People (and Firms) Value Tax Privacy?

Many people and firms express concerns about tax privacy. According to the 2022 Taxpayer Experience Survey conducted by the Internal Revenue Service (2023), 18 percent of taxpayers trust the IRS "very much" to keep their information safe, 54 percent trust it to do so "somewhat," 20 percent trust it "not very much," and 8 percent trust it "not at all." (Of course, some people might believe that the IRS keeps their personal my information "safe" while at the same time believing that is has no right to have the information.) A 2023 Pew survey found that 71 percent of people are concerned about how government uses the data it collects about them, compared to 81 percent for companies (McClain et al. 2023). At the same time, 77 percent say they have little to no understanding about what the government does with the data it collects about them (up from 64 percent in 2019), compared to 67 percent for companies. These attitudes are correlated with political affiliation: the 71 percent figure is 77 percent for Republicans, and 65 percent for Democrats.

A number of empirical studies have attempted to quantify individual valuations of privacy in diverse non-tax contexts—such as the value people place on personal information revealed online (Hann et al. 2007), online borrowers subject to disclosure requirements (Tang 2019), or removal from marketers' call lists (Varian, Wallenberg, and Woroch 2005). Miller and Tucker (2009) measure how much state privacy regulations affected the adoption of electronic medical records by hospitals, and Demirer et al. (2024) measure how much the European Union's General Data Protection Regulation rules on data storage and data processing affected data storage and processing by EU firms relative to comparable US firms.[5] However, the value that people place on privacy may be highly context-dependent. In a field

[5] Much recent non-tax research about privacy is collected in Goldfarb and Tucker (2024).

experiment, Acquisti, John, and Loewenstein (2013) show that the implicit valuations of privacy are highly context-dependent, as well as affected by endowment and the order in which different privacy options are described. However, very little research attempts to measure the value people place specifically on tax privacy.

In survey-based measures of the valuation of privacy, a common pattern is that many individuals perceive a high cost to information disclosure, but also that they are only willing to spend a small fraction of that cost to protect privacy. Thus, the value they provide for willingness-to-accept less privacy is much higher than their willingness-to-pay for protecting privacy. In the Acquisti, John, and Loewenstein (2013) field experiment, the ratio of willingness-to-accept relative to willingness to pay is 5.47; for comparison, the average ratio observable for ordinary private goods is 2.92 (as reported in Horowitz and McConnell 2002). In a preliminary survey, Sanz-Maldonado (2023) finds that the median willingness-to-pay to protect last year's tax and income from public disclosure is (just) $50. However, due to the presence of a long tail, the average willingness-to-pay is $57 and the average willingness-to-accept is $3,491. The average willingness-to-accept is 13.5 times larger than the average willingness-to-pay, and the median willingness-to-accept is four times larger than the median willingness-to-pay.

One can also learn about the value people place on tax privacy by observing how people and firms react to public disclosure programs. In an analysis of data from US nonprofit entities, Karol and Sanz-Maldonado (2024) take advantage of the fact that many of these entities are required to publicly disclose the compensation of certain employees whose earnings exceed $100,000. Bunching of reported earnings below this cutoff provides information about the perceived net cost, to employers and employees, of having one's compensation in the public domain.

As another example, Japan had a public disclosure system for taxes of high-income individuals until 2004. The disclosure applied only to individuals and firms with taxable income above ¥40 million (about $400,000). Hasegawa et al. (2013) found strong evidence based on the bunching of observations right below the disclosure threshold. For businesses, this behavior was consistent with the local characterization of "39 companies," whose reported taxable income was kept below the disclosure threshold of ¥40 million so as not to provide evidence about their profitability, which could have adversely affected the deals they could make with other companies.

Similarly, Australia requires publicly disclosing tax information about large companies with over $100 million (in Australian dollars) of "total income" on their tax return. Hoopes, Robinson, and Slemrod (2018) find that in the first year that the disclosure began, for foreign-owned, Australian private, and Australian public companies, there was a jump in the number of firms reporting (to the tax authority) an amount just under the threshold, while the number reporting income just over the threshold held relatively steady or declined. The excess mass just under the disclosure threshold was relatively less pronounced for public firms, with an increase of just 33 percent in the number of firms from 2013 to 2014. In contrast, the number of firms just under the threshold increased by 65 percent and 73 percent for private and foreign-owned firms, respectively. This finding is consistent with the argument that, for private companies, the costs of disclosure are relatively high, because less information is already in the public domain.

Using a similar empirical research design but based on individual-level data, Pitt and Slemrod (1989) and Benzarti (2020) infer the extent of compliance costs incurred in the process of itemizing deductions for income tax purposes by estimating how much tax saving taxpayers forego by taking the standard deduction rather than itemizing deductions. But itemizing deductions requires revealing to the Internal Revenue Service potentially sensitive information about charitable contributions, medical expenses, and other expenditures. Thus, this study and the earlier studies are likely generating an estimate of the sum of compliance costs and the privacy costs of itemizing, rather than of compliance costs alone; disentangling the two awaits further research attention.

This possible confusion between compliance costs and privacy considerations also applies to the recent literature that explores the impact of changes in information reporting requirements on taxpayer behavior. For example, Garbinti et al. (2023) investigate the impact of a simplification in the reporting requirements of the French wealth tax. They find sharp bunching at information requirement discontinuities, but not at "pure" tax rate kinks; they provide evidence suggesting that privacy concerns were not an important factor in the behavioral response. Fack and Landais (2016) study the effect of tightening the reporting requirements for claiming charitable contributions in France. They find a substantial drop in reported contributions, which they attribute to reduced over-reporting, but do not investigate the possible role of privacy concerns.

To evaluate the benefits and costs of altering tax privacy, it is essential to have a sense of what value taxpayers place on the government having access to certain information. While some progress has been made from surveys and through inferring the value of privacy from observing behavior, this remains an open research question.

Privacy and Tax Evasion

To this point I have proceeded as if the information that the government collects for tax purposes always represents the unalloyed truth, which is decidedly not correct. The US Internal Revenue Service (2023) estimated that 14 percent of tax liability is not remitted. Moreover, the IRS methodology for estimating the underpayment in taxes owed may underestimate the tax noncompliance of the top 1 percent of the income distribution by as much as 50 percent (Guyton et al. 2021).

A taxpayer who objects to the government possessing accurate personal information might pass along false information—when that reduces tax liability, we call it tax evasion. Greater privacy rights will also open opportunities for greater tax evasion. This suggests a number of interesting connections between privacy and evasion.

First, the trust that citizens have for the government to protect their privacy may affect the quality of the information that taxpayers are willing to share with the government. Trust in government may then affect tax compliance (Luttmer and Singhal 2014; OECD 2019). It has been claimed that, at one time in Colombia, affluent people understated their wealth to the tax authority not primarily to save on taxes, but rather to reduce the chance they would be kidnapped and held for ransom (Londoño-Vélez 2012, p. 48). Recall that public disclosure in Japan was abolished in part because it was thought that it led to harassment of taxpayers disclosed to be affluent.

Second, public disclosure of tax information might reduce evasion, at least of the self-employed, given the possibility that the now-public information could be passed along to the tax authorities by neighbors or business competitors. Analysis of the Norwegian and Pakistani tax disclosure episodes (mentioned earlier) suggests that it does modestly reduce evasion among self-employed individuals, while studies of the Australian and Japanese policies find no evidence that corporate tax compliance was affected (Bø, Slemrod, and Thoresen 2015; Hasegawa et al. 2013; Hoopes, Robinson, and Slemrod 2018; Slemrod, Ur Rehman, and Waseem 2022).

Public debates about whether to increase the power of tax authorities often collide with arguments about privacy. In 2021, for example, the Biden administration proposed that Internal Revenue Service should have power to monitor annual deposits and withdrawals from accounts at banks and other financial institutions, for transfers exceeding $600 within accounts exceeding $10,000. A letter to Congressional leaders from an industry group exemplified the objection on privacy grounds: "Indiscriminate, blanket data collection would amount to a troubling effort to profile American taxpayers based on account characteristics without grounds for suspicion of tax evasion" (SBE Council 2021).

Similarly, arguments about loss of privacy and intrusiveness dominated criticism of the provision of $80 billion of funds (later cut to $60 billion) over a decade provided to the Internal Revenue Service by the Inflation Reduction Act of 2022. Senator Lindsey Graham (R-SC) offered this objection: "If you think the federal government is out of control now, God help us when you get 87,000 new IRS agents who are looking under every rock and stone to get money out of your pocket" (as quoted in Burns 2022).

As yet another example, the Internal Revenue Service announced a free tax filing in October 2023, to be piloted in 13 states for the 2024 filing season. Opponents claimed it was a waste of money addressing a second-order tax issue, but also pointed to the encroachment on taxpayer privacy. Grover Norquist, president of Americans for Tax Reform, said: "'If the government does your taxes for you, they have to know everything about you. It is the end of economic privacy" (as quoted in Rappeport 2023).

Of course, objections to increased tax enforcement can arise from a variety of motives, including a fear that it will lead to higher taxes owed. But the fact that the language of privacy is often used to express these objections suggests a belief that privacy concerns are salient for many taxpayers.

Offering Choices to Taxpayers about Information Revelation

How might tax policy be designed to reduce or minimize the extent of personal information revealed? As mentioned earlier, one possibility would be to replace the income tax with a commodity tax system that levied tax rates based on goods' income elasticity, taxing luxury goods at a relatively high rate and necessary goods at a zero, or even negative, rate.[6] Such a tax would generate more inter-good distortion than an income tax that, under some fairly weak assumptions, implies a higher efficiency cost, and would cause some horizontal inequity based on heterogeneous preferences—the yacht-loving high-income household would bear a higher burden than the affluent household with pedestrian tastes. Also, in practical terms, replacing the income tax with a revenue-equivalent differentiated commodity tax would require the political system to adopt relatively high rates on many goods and services.

An alternative approach would be to give taxpayers options about what information they would prefer to reveal, and to adjust taxes accordingly. As a conceptual example, consider conditioning personal tax liability on "tags"—information that correlates with ability or income and is completely or nearly inalterable. Akerlof (1978) argued that, even if one accepts that such characteristics are not, in and of themselves, relevant for redistribution, they still have a role to play in designing optimal tax because their use can reduce the social cost of income redistribution based on alterable measures such as income. The most commonly raised potential tags are race, height, age, and gender; Logue and Slemrod (2008) also consider genetic tags. To assess whether tags should be part of a tax system, one should weigh the social welfare gains against the value of the privacy lost when these characteristics are revealed to the government, as well as the distributional implications.

In a political setting, the most likely form of tax tags would be to offer subsidies on the basis of certain "negative" tags that are correlated with low income, rather than having "positive" markers trigger additional tax liability, as Logue and Slemrod (2008) suggest. In this case, a taxpayer who prefers could opt out, not provide the negative information, and forego the subsidy. This could apply to most commonly discussed possible tags, such as (less-than-average) height, (minority) race, and (female) gender.[7] Of course, people will not want to reveal tags that increase tax liability, but that is mitigated by the incentive of people with "low" tags to reveal

[6] A linear tax on labor income plus differential excise taxes could get part of the way there. For employees, this can be implemented by employers remitting the appropriate tax liability, without knowing anything else about the employee other than the compensation paid and the government not knowing anything at all, except when enforcement action is needed. This would be exact withholding, in that no refund or tax exists at the close of the tax year. But this system is not likely to work well for self-employed people and for employees in small businesses where collusion between employer and employees is facilitated.

[7] Some, including Brown (2021), and Bearer-Friend (2022), have advocated for the IRS to ask for, or otherwise obtain, taxpayers' race (it does not now), on the grounds that it would facilitate analysis and publication of the impact of the tax system by race. Having data on taxpayer race would also presumably facilitate using the tax code to implement race-based reparations that some, including Brown (2021), advocate.

their information, allowing inferences to be made about the average setting of unrevealed tags.

More generally, the tax system could perhaps make greater use of rules for revelation of information that have developed in the online context. I have in mind both opt-in rules, where consent requires that users actively and explicitly agree to data collection, processing, or sharing; and opt-out rules, where user consent is assumed unless users act to withdraw it. Although the same nomenclature is not applied, many essentially opt-in and opt-out features exist in tax systems. For example, if one does not wish to divulge one's charitable contributions or medical expenses, one can simply not report them, and thereby forego the potential tax saving from claiming the deduction. It seems worthwhile for economists and policymakers to pay more attention to these kinds of opt-in/opt-out rules, which offer taxpayers an option for greater privacy, but at a cost, thus inducing taxpayers to reveal their true valuation of government access to personal information.

Privacy and Wealth Taxes

The United States had a national debate about enacting a wealth tax during the 2020 election, with one prominent proposal defended by Saez and Zucman (2019). Such a tax typically features a relatively high exemption level and a very broad measure of taxable wealth. The wealth tax controversy is still alive in several states (Amer 2022; Picchi 2023).

Opponents of a wealth tax in the United States focused on the expansion of IRS authority it would require, often lumping together the purported invasion of privacy and the additional compliance and administrative costs it would generate. One journalist wrote: "[Monitoring] household wealth . . . will require extensive investigative power on the part of the government. What's that antique car worth in your garage? What is the revenue growth rate of your company? How did you calculate the value of your lakeside cottage now that you put the new boat dock in?" (Willey 2023).[8]

Here, I set aside the full range of arguments over the merits of a wealth tax and focus on the privacy issues. A wealth tax would not encroach on the privacy of what many consider the most sensitive items of personal information, such as sexual preference, health status, or spending habits. Indeed, some objections to a wealth tax on the grounds of privacy encroachment might be a smokescreen for the real objection of increased tax liability. A helpful thought experiment is to contemplate the debate if the information needed for a wealth tax was collected only for statistical purposes

[8] Such arguments have been made to counter other new taxes. When the state of Michigan announced it was considering road usage taxes to replace lost revenue from the gas tax due to increased electronic vehicle use, a critic argued that there was no way to levy such a tax "without a significant growth of government intrusion into our privacy" (as quoted in Gibbons 2024).

(say, through the triennial Federal Reserve Survey of Consumer Finances), with no attendant additional tax (or regulatory) liability for anyone.

However, a wealth tax would require disclosing to the government a considerable amount of information beyond what is now collected to administer the income tax; for example, current values of financial assets, real estate, private businesses, cars, and trust funds. Much of the lively academic back-and-forth about the information requirements of a wealth tax—for example, Saez and Zucman (2019) versus Kopczuk (2019)—focused on the administrative and compliance costs of wealth measurement.

Some of the discussion over wealth taxes nicely highlights the difference between collection costs and privacy. Consider the need to know the value of financial assets, given that the capital income flow is already often reported by the financial institution, and the value is known to the institution. In this setting, the collection costs of the government obtaining this information would likely be low. But the government would now have access to information it did not have before, and in some cases could not easily infer. When presumptive methods of valuation are used to determine asset values, as the Swiss do for their wealth tax by basing the taxable value of private businesses on, among other things, profits, then the privacy of business profits is compromised.

Conclusion

The strength of the connection between optimal tax policy and privacy depends on the valuation that citizens place on providing access to particular information or on keeping that information private. This article has focused on the current state of economic analysis of privacy concerns and tax policy, but has sidestepped some closely related issues.

First, there is inter-country competition for privacy, like competition regarding tax (lack of) enforcement. International data sharing agreements, such as the Common Reporting Standard, involve financial institutions sharing information to reduce the opportunities for tax evasion across national borders. Avi-Yonah and Mazzoni (2016, p. 2) discuss tensions such as automatic exchange of information and country-by-country reporting and upholding human rights such as privacy. They argue that "within the process of international exchange of information, individuals have only procedural safeguards, not a substantive right to privacy." Furthermore, international agreements may have to consider that different countries have different views and standards about privacy.

A second issue is the possibilities for privacy created by cryptocurrency. Proponents of cryptocurrency use often argue that the pseudonymous nature of these digital assets protects privacy and allows individuals to conduct transactions outside the traditional banking system and taxation authorities, while critics see this as facilitating tax evasion and illicit activities.

Finally, citizen demand for privacy not only affects the tax system but also the availability of tax micro data to outside researchers, as well as the rules about tax data sharing within the government. Useful starting points on these issues include Domingo-Ferrer and Muralidhar (2020), Aziz et al. (2023), Burman et al. (2024), and the two-paper symposium on privacy protection and government data in this journal with papers by Bowen (2024) and Ruggles (2024). I suspect, but cannot (yet) prove, that there is a (negative) correlation across countries between citizen concern for privacy and the availability of data to outside researchers. Economic researchers need to ponder the implications of the correlations among privacy-mediated data availability, data quality, and key tax system parameters such as the elasticity of tax bases. As a small step in this direction, Hsieh, Sanz-Maldonado, and Slemrod (2024) investigate the relationship between citizen concern for privacy and the elasticity of taxable income, as limiting government access to taxpayer information constrains tax system measures that would reduce this elasticity.

Back in 2006, I wrote that the "revolution in information technology—in storing, processing, and moving information—has reduced the cost of control and the cost of invading privacy" (Slemrod 2006). I expressed concern that productivity growth in information technology might disadvantage the government through a Baumolian "cost disease" (Baumol and Bowen 1965), not because the government, like the performing arts, is inherently less able to profit from technology-related productivity gains, but rather that, because of privacy concerns, the citizenry does not want the government to take advantage of these gains. The ongoing information revolution has made these concerns—in tax policy and indeed across the policy spectrum—even more salient over time.

■ *I thank Gerardo Sanz-Maldonado for helpful conversations, Youssef Benzarti, Joshua Blank, Deirdre James, Marjorie Kornhauser, and participants in the NBER Trans-Atlantic Public Economics Seminar (TAPES) on Tax and Transfer Systems in Helsinki, June 12–14, 2024, for valuable comments on an earlier draft, and Noah Rich for able research assistance. Thanks also to the JEP editorial team—Heidi Williams, Nina Pavcnik, Jonathan Parker, and Timothy Taylor—for feedback and suggestions.*

References

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein. 2015. "Privacy and Human Behavior in the Age of Information." *Science* 347 (6221): 509–14.

Acquisti, Alessandro, Leslie K. John, and George Loewenstein. 2013. "What is Privacy Worth?" *Journal of Legal Studies* 42 (2): 249–74.

Akerlof, George A. 1978. "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." *American Economic Review* 68 (1): 8–19.

Amer, Yasmin. 2022. "Massachusetts Voters Approve 'Millionaires Tax' Ballot Question." *WBUR*, November 9. https://www.wbur.org/news/2022/11/09/massachusetts-election-question-1-tax.

Auxier, Brooke, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. 2019. "Americans and Privacy: Concerned, Confused, and Feeling Lack of Control over Their Personal Information." Pew Research Center, November 15. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

Avi-Yonah, Reuven S., and Gianluca Mazzoni. 2016. "Taxation and Human Rights: A Delicate Balance." University of Michigan Public Law Research Paper 520.

Aziz, Rezak, Soumya Banerjee, Samia Bouzefrane, and Thinh Le Vinh. 2023. "Exploring Homomorphic Encryption and Differential Privacy Techniques towards Secure Federated Learning Paradigm." *Future Internet* 15 (9): 310.

Baumol, William J., and William G. Bowen. 1965. "On the Performing Arts: The Anatomy of their Economic Problems." *American Economic Review* 55 (1/2): 495–502.

Bearer-Friend, Jeremy. 2022. "Colorblind Tax Enforcement." *New York University Law Review* 97 (1): 1–58.

Benzarti, Youssef. 2020. "How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs." *American Economic Journal: Economic Policy* 12 (4): 38–57.

Blank, Joshua D. 2011. "In Defense of Individual Tax Privacy." *Emory Law Journal* 61: 265–348.

Bø, Erlend E., Joel Slemrod, and Thor O. Thoresen. 2015. "Taxes on the Internet: Deterrence Effects of Public Disclosure." *American Economic Journal: Economic Policy* 7 (1): 36–62.

Bowen, Claire McKay. 2024. "Government Data of the People, by the People, for the People: Navigating Citizen Privacy Concerns." *Journal of Economic Perspectives* 38 (2): 181–200.

Brown, Dorothy A. 2021. *The Whiteness of Wealth: How the Tax System Impoverishes Black Americans—And How We Can Fix It.* Crown.

Burman, Leonard, Barry Johnson, Victoria L. Bryant, Graham MacDonald, and Robert McClelland. 2024. "Protecting Privacy and Expanding Access in a Modern Administrative Tax Data System." *National Tax Journal* 77 (4).

Burns, Tobias. 2022. "GOP Rails against IRS Funding in Inflation Reduction Act." *The Hill (blog)*, August 10. https://thehill.com/policy/3594879-gop-rails-against-irs-funding-in-inflation-reduction-act/.

Dekel, Inbal, Rachel Cummings, Ori Heffetz, and Katrina Ligett. 2022. "The Privacy Elasticity of Behavior: Conceptualization and Application." NBER Working Paper 30215.

Demirer, Mert, Diego J. Jiménez Hernández, Dean Li, and Sida Peng. 2024. "Data, Privacy Laws and Firm Production: Evidence from the GDPR." NBER Working Paper 32146.

Domingo-Ferrer, Josep, and Krishnamurty Muralidhar, eds. 2020. *Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2020, Tarragona, Spain, September 23–25, 2020, Proceedings.* Springer Nature.

Eisinger, Jesse, Jeff Ernsthausen, and Paul Kiel. 2021. "The Secret IRS Files: Trove of Never-Before-Seen Records Reveal How the Wealthiest Avoid Income Tax." *ProPublica*, June 8. https://www.propublica.org/article/the-secret-irs-files-trove-of-never-before-seen-records-reveal-how-the-wealthiest-avoid-income-tax.

Fack, Gabrielle, and Camille Landais. 2016. "The Effect of Tax Enforcement on Tax Elasticities: Evidence from Charitable Contributions in France." *Journal of Public Economics* 133: 23–40.

Fulda, Joseph S. 1996. "Liberty and Privacy: Connections." Foundation for Economic Education, December 1. https://fee.org/articles/liberty-and-privacy-connections/#:~:text=If%20property%20is%20liberty's%20other,can%20possibly%20interfere%20with%20it.

Garbinti, Bertrand, Jonathan Goupille-Lebret, Mathilde Muñoz, Stefanie Stantcheva, and Gabriel Zucman. 2023. "Tax Design, Information, and Elasticities: Evidence from the French Wealth Tax." NBER Working Paper 31333.

Garriga, Santiago, and Dario Tortarolo. 2024. "Wage Effects of Means-Tested Transfers: Incidence Implications of Using Firms as Intermediaries." Institute for Fiscal Studies Working Paper 24/29.

Gibbons, Lauren. 2024. "A mileage tax for Michigan motorists? Lawmakers want $5m for pilot program." Bridge Michigan, May 19. https://www.bridgemi.com/michigan-government/mileage-tax-michigan-motorists-lawmakers-want-5m-pilot-program.

Goldfarb, Avi, and Catherine E. Tucker, eds. 2024. *The Economics of Privacy.* University of Chicago Press.

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.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong Tom Lee, and Ivan P. L. Png. 2007. "Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach." *Journal of Management*

Information Systems 24 (2): 13–42.
Hasegawa, Makoto, Jeffrey L. Hoopes, Ryo Ishida, and Joel Slemrod. 2013. "The Effect of Public Disclosure on Reported Taxable Income: Evidence from Individuals and Corporations in Japan." *National Tax Journal* 66 (3): 571–607.
Hoopes, Jeffrey L., Leslie Robinson, and Joel Slemrod. 2018. "Public Tax-Return Disclosure." *Journal of Accounting and Economics* 66 (1): 142–62.
Hoopes, Jeffrey L., Leslie Robinson, and Joel Slemrod. Forthcoming. "Corporate Tax Disclosure." Journal of the American Taxation Association.
Horowitz, John K., and Kenneth E. McConnell. 2002. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management* 44 (3): 426–47.
Hsieh, Meng Hsuan, Gerardo Sanz-Maldonado, and Joel Slemrod. 2024. "External Validity in Empirical Public Finance." Unpublished.
Karol, Stephanie, and Gerardo Sanz-Maldonado. 2024. "Salary Disclosure and the Value of Tax Privacy: Evidence from U.S. Nonprofits." Unpublished.
Keen, Michael, and Joel Slemrod. 2021. *Rebellion, Rascals, and Revenue: Tax Follies and Wisdom throug the Ages*. Princeton University Press.
Kopczuk, Wojiech. 2019. "Comments on 'Progressive Wealth Taxation' by Emmanuel Saez and Gabriel Zucman." *Brookings Papers on Economic Activity* 50 (2): 512–26.
Kopczuk, Wojciech, Justin Marion, Erich Muehlegger, and Joel Slemrod. 2016. "Does Tax-Collection Invariance Hold? Evasion and the Pass-Through of State Diesel Taxes." *American Economic Journal: Economic Policy* 8 (2): 251–86.
Kornhauser, Marjorie E. 2010. "Shaping Public Opinion and the Law: How a 'Common Man' Campaign Ended a Rich Man's Law." *Law and Contemporary Problems* 73 (1): 123–47.
Kornhauser, Marjorie E. 2018. "The Invisible Government and Conservative Tax Lobbying 1935–1936." *Law and Contemporary Problems* 81 (2): 167–201.
Krähmer, Daniel, and Roland Strausz. 2023. "Optimal Nonlinear Pricing with Data-Sensitive Consumers." *American Economic Journal: Microeconomics* 15 (2): 80–108.
Langer, Emily, 2014. "Johnnie Walters, IRS Commissioner under President Richard M. Nixon, Dies at 94." *Washington Post,* June 26. https://www.washingtonpost.com/national/johnnie-walters-irs-commissioner-under-president-richard-m-nixon-dies-at-94/2014/06/26/e6ae4906-fd3d-11e3-b1f4-8e77c632c07b_story.html.
Leff, Mark H. 1984. *The Limits of Symbolic Reform: The New Deal and Taxation, 1933–1939*. Cambridge University Press.
Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. Basic Books.
Logue, Kyle, and Joel Slemrod. 2008. "Genes as Tags: The Tax Implications of Widely Available Genetic Information." *National Tax Journal* 61 (4.2): 843–63.
Londoño-Vélez, Juliana. 2012. "Income and Wealth at the Top in Colombia: An Exploration of Tax Records 1993–2010." Master's dissertation, Paris School of Economics.
Luttmer, Erzo F. P., and Monica Singhal. 2014. "Tax Morale." *Journal of Economic Perspectives* 28 (4): 149–68.
McClain, Colleen, Michelle Faverio, Monica Anderson, and Eugenie Park. 2023. "How Americans View Data Privacy." Pew Research Center, October 18. https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/.
Miller, Amalia R., and Catherine Tucker. 2009. "Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records." *Management Science* 55 (7): 1077–93.
Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.
Nakayama, Kiyoshi. 2021. *Public Disclosure of Taxpayer Information as a Measure to Improve Tax Compliance*. IMF.
Neate, Rupert. 2018. "Over $119bn Wiped off Facebook's Market Cap after Growth Shock" *Guardian*, July 26. https://www.theguardian.com/technology/2018/jul/26/facebook-market-cap-falls-109bn-dollars-after-growth-shock.
OECD. 2019. *Tax Morale: What Drives People and Businesses to Pay Tax?* OECD Publishing.
Parets, Michael. 2023. "How Tax Transparency Affects Privacy for High Net Worth Individuals." EY, October 12. https://www.ey.com/en_tw/tax/how-tax-transparency-affects-privacy-for-high-net-worth-individuals.
Perez-Truglia, Ricardo. 2020. "The Effects of Income Transparency on Well-Being: Evidence from a

Natural Experiment." *American Economic Review* 110 (4): 1019–54.

Picchi, Aimee. 2023. "A National Wealth Tax Has Gone Nowhere. Now Some States Want to Tax the Ultra-rich." *CBS News*. January 20. https://www.cbsnews.com/news/wealth-tax-in-8-states-california-new-york-connecticut-washington-illinois/.

Pitt, Mark M., and Joel Slemrod. 1989. "The Compliance Cost of Itemizing Deductions: Evidence from Individual Returns." *American Economic Review* 79 (5): 1224–32.

Pourbabaee, Farzad, and Federico Echenique. 2024. "Binary Mechanisms under Privacy-Preserving Noise." Preprint, arXiv. https://doi.org/10.48550/arXiv.2301.06967.

Purcell, Jay, and Vana Rossi. 2019. "Privacy vs. Transparency." *IMF Finance and Development*, September. https://www.imf.org/en/Publications/fandd/issues/2019/09/illicit-financial-flows-and-privacy-vs-transparency-purcell.

Rappeport, Alan. 2023. "Lobbyists Begin Chipping Away at Biden's $80 Billion IRS Overhaul." *New York Times*, March 30. https://www.nytimes.com/2023/03/30/us/politics/biden-irs-overhaul-taxes.html.

Reck, Daniel, Joel Slemrod, and Trine Engh Vattø. 2022. "Public Disclosure of Tax Information: Compliance Tool or Social Network?" *Journal of Public Economics* 212: 104708.

Ruggles, Steven. 2024. "When Privacy Protection Goes Wrong: How and Why the 2020 Census Confidentiality Program Failed." *Journal of Economic Perspectives* 38 (2): 201–26.

Saez, Emmanuel, and Gabriel Zucman. 2019. "Progressive Wealth Taxation." *Brookings Papers on Economic Activity* 50 (2): 437–511.

Sanz-Maldonado, Gerardo. 2023. "Using a Survey to Value Tax Privacy." Unpublished.

SBE Council. 2021. "Joint Letter to Congress Opposing Consumer Financial Accounting Reporting." September 13. https://sbecouncil.org/2021/09/13/joint-letter-to-congress-opposing-consumer-financial-accounting-reporting/.

Schmidt, Michael S. 2022. "Trump Wanted I.R.S. Investigations of Foes, Top Aide Says." *New York Times*, November 13. https://www.nytimes.com/2022/11/13/us/politics/trump-irs-investigations.html.

Slemrod, Joel. 2006. "Taxation and Big Brother: Information, Personalisation and Privacy in 21st Century Tax Policy." *Fiscal Studies* 27 (1): 1–15.

Slemrod, Joel. 2008. "Does It Matter Who Writes the Check to the Government? The Economics of Tax Remittance." *National Tax Journal* 61 (2): 251–75.

Slemrod, Joel, Obeid Ur Rehman, and Mazhar Waseem. 2022. "How Do Taxpayers Respond to Public Disclosure and Social Recognition Programs? Evidence from Pakistan." *Review of Economics and Statistics* 104 (1): 116–32.

Smith, Adam. 1776. *The Wealth of Nations*. University of Chicago Press, 1976.

Stuntz, William J. 1995. "Privacy's Problem and the Law of Criminal Procedure." *Michigan Law Review* 93 (5): 1016–78.

Tang, Huan. 2019. "The Value of Privacy: Evidence from Online Borrowers." Unpublished.

Treasury Inspector General for Tax Administration. 2017. *Review of Selected Criteria Used to Identify Tax-Exempt Applications for Review*. Treasury Inspector General for Tax Administration.

US Department of Justice, Office of Public Affairs. 2017. "Attorney General Jeff Sessions Announces Department of Justice has Settled with Plaintiff Groups Improperly Targeted by IRS." Press release, October 26. https://www.justice.gov/opa/pr/attorney-general-jeff-sessions-announces-department-justice-has-settled-plaintiff-groups.

US Internal Revenue Service. 2023. "Tax Gap Projections for Tax Year 2020." IRS Publication 5870. https://www.irs.gov/pub/irs-pdf/p5870.pdf.

US Internal Revenue Service. 2023. *IRS Report to Congress: Inflation Reduction Act §10301(1)(B) IRS-run Direct e-File Tax Return System*. US Department of the Treasury.

Varian, Hal R. 2009. "Economic Aspects of Personal Privacy." In *Internet Policy and Economics: Challenges and Perspectives*, edited by William H. Lehr and Lorenzo Maria Pupillo, 101–09.

Varian, Hal R., Fredrik Wallenberg, and Glenn Woroch. 2005. "The Demographics of the Do-Not-Call List." *IEEE Security and Privacy* 3 (1): 34–39.

Willey, Bruce. 2023. "Why a Wealth Tax Would Be Terrible for American Taxpayers." Kiplinger, October 4. https://www.kiplinger.com/taxes/why-a-wealth-tax-would-be-terrible-for-american-taxpayers.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 225–246

Philipp Strack, 2024 Clark Medalist

Drew Fudenberg

Philipp Strack is creative, insightful, and skillful, which has allowed him to make major contributions to many areas of microeconomic theory, including behavioral economics, information acquisition and learning, and mechanism design. Some of these papers provide a new understanding of important economics phenomena, others introduce results and techniques that will be used for years to come, and some of the papers do both. Together they have helped spark what his Clark Medal citation called a "new wave of information economics."

Philipp was born in Bonn, Germany, to a Greek mother (Nora Andrikopoulou, an archeologist) and a German father (Ulrich Strack, a lawyer). In high school, he took college-level math classes at the University of Bonn, where he subsequently matriculated. In college, Philipp was greatly influenced by Paul Heidhues and Benny Moldovanu, and he became fascinated by the idea that abstract reasoning and mathematical models can used to study social questions. This led him to earn *diploms* (roughly equivalent to a master's degree) at Bonn in both economics and mathematics and then start PhD work in both disciplines simultaneously. Luckily for the economics profession, he subsequently decided to focus on economics and completed his economics PhD in 2013. At that time, new PhDs from Bonn and other European schools were often overlooked by economics departments in the United States, and my department only became aware of him relatively late in the process. In the end, we did have him come for a flyout. I was very impressed by his intellect and his drive, and Tomasz Strzalecki and I invited him to work with us on our drift-diffusion project.

■ *Drew Fudenberg is the Paul A. Samuelson Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20241422.

Philipp Strack

Philipp started his career at the University of California, Berkeley, where he was promoted to tenure before moving to Yale University in 2019. In 2023, Philipp won the Bodossaki Distinguished Young Scientist Award, which is given every two years to young Greek scientists whose "outstanding performance has already significantly contributed to the advancement of science." As I will show, this accolade was well-deserved, as was the Clark Medal that has now followed. I will refer to his papers mentioned in this essay by number, as listed in Table 1.

Mechanism Design

Mechanism design and its applications have been one of the great successes of modern economic theory, with many real-world applications including auction design and allocating medical school students to residencies. The initial wave of work in this area tended to focus on one-time settings and "rational" agents with standard expected utility preferences. However, many settings are dynamic, because the players may encounter the same mechanism more than once, or may observe how the mechanism has worked out in previous iterations. Moreover, in some cases, non-expected-utility preferences that allow, for example, loss aversion are a better description of the players' preferences and behavior. Philipp's work has provided useful new tools for mechanism design and applied them to extend the analysis of mechanism design both in classic settings and some new ones.

Table 1
Selected Papers of Philipp Strack

1. “Gambling in contests” (with Christian Seel). 2013. *Journal of Economic Theory* 148: 2033–48.
2. “Dynamic Revenue Maximization: A Continuous Time Approach” (with Dirk Bergemann). 2015. *Journal of Economic Theory* 159: 819–53.
3. “Until the Bitter End: On Prospect Theory in a Dynamic Context” (with Sebastian Ebert). 2015. *American Economic Review* 105: 1618–33.
4. “Optimal Stopping with Private Information” (with Thomas Kruse). 2015. *Journal of Economic Theory* 159: 702–27.
5. “Active learning with a misspecified prior” (with Drew Fudenberg and Gleb Romanyuk). 2017. *Theoretical Economics* 12: 1155–89.
6. “Never, ever getting started: On prospect theory without commitment” (with Sebastian Ebert). 2018. Working Paper.
7. “Revenue-Maximizing Mechanisms with Strategic Customers and Unknown, Markovian Demand” (with Alex Gershkov and Benny Moldovanu). 2018. *Management Science* 64: 2031–46.
8. “Speed, Accuracy, and the Optimal Timing of Choices” (with Drew Fudenberg and Tomasz Strzalecki). 2018. *American Economic Review* 108: 3651–84.
9. “Unrealistic Expectations and Misguided Learning” (with Paul Heidhues and Botond Köszegi). 2018. *Econometrica* 86: 1159–214.
10. “Identifying procrastination from the timing of choices” (with Paul Heidhues). 2019. Tech. rep. Working Paper.
11. “Bitcoin: An Axiomatic Approach and an Impossibility Theorem” (with Jacob D. Leshno). 2020. *American Economic Review: Insights* 2: 269–286.
12. “Optimal control of an epidemic through social distancing” (with Thomas Kruse). 2020. Available at SSRN 3581295.
13. “Stochastic Dominance under Independent Noise” (with Luciano Pomatto and Omer Tamuz). 2020. *Journal of Political Economy* 128: 1877–1900.
14. “Testing the drift-diffusion model” (with Drew Fudenberg, Whitney Newey, and Tomasz Strzalecki). 2020. *Proceedings of the National Academy of Sciences* 117: 33141–48.
15. “Turning up the heat: The discouraging effect of competition in contests” (with Dawei Fang and Thomas Noe). 2020. *Journal of Political Economy* 128: 1940–75.
16. “A theory of Auctions with Endogenous Valuations” (with Alex Gershkov, Benny Moldovanu, and Mengxi Zhang). 2021. *Journal of Political Economy* 129: 1011–51.
17. “Convergence in models of misspecified learning” (with Paul Heidhues and Botond Köszegi). 2021. *Theoretical Economics* 16: 73–99.
18. “Extreme Points and Majorization: Economic Applications” (with Andreas Kleiner and Benny Moldovan). 2021. *Econometrica* 89: 1557–93.
19. “From Blackwell Dominance in Large Samples to Rényi Divergences and Back Again” (with Xiaosheng Mu, Luciano Pomatto, and Omer Tamuz). 2021. *Econometrica* 89: 475–506.
20. “Limit Points of Endogenous Misspecified Learning” (with Drew Fudenberg and Giacomo Lanzani). 2021. *Econometrica* 89: 1065–98.
21. “Rational Groupthink” (with Matan Harel, Elchanan Mossel, and Omer Tamuz). 2021. *Quarterly Journal of Economics* 136: 621–68.
22. “Optimal Auctions: Non-expected Utility and Constant Risk Aversion” (with Alex Gershkov, Benny Moldovanu, and Mengxi Zhang). 2022. *Review of Economic Studies* 89: 2630–62.

(*continued*)

Table 1
Selected Papers of Philipp Strack (continued)

23. "Equity in Dynamic Matching: Extreme Waitlist Policies" (with Afshin Nikzad). 2023. *Management Science*.
24. "Optimal Insurance: Dual Utility, Random Losses, and Adverse Selection" (with Alex Gershkov, Benny Moldovanu, and Mengxi Zhang). 2023. *American Economic Review* 113: 2581–614.
25. "Pathwise Concentration Bounds for Misspecified Bayesian Beliefs" (with Drew Fudenberg and Giacomo Lanzani). 2023. *Theoretical Economics* 18: 1585–622.
26. "The Cost of Information: The Case of Constant Marginal Costs" (with Luciano Pomatto and Omer Tamuz). 2023. *American Economic Review* 113: 1360–93.
27. "Background Risk and Small-Stakes Risk Aversion" (with Xiaosheng Mu , Luciano Pomatto, and Omer Tamuz). 2024a. *American Economic Review: Insights* 6: 262–76.
28. "Learning in Repeated Interactions on Networks" (with Wanying Huang and Omer Tamuz). 2024. *Econometrica* 92: 1–27.
29. "Monotone Additive Statistics" (with Xiaosheng Mu , Luciano Pomatto, Philipp Strack, and Omer Tamuz). 2024b. *Econometrica* 94: 995–1031.
30. "Privacy Preserving Signals" (with Kai Hao Yang). 2024. Working paper.
31. "Selective Memory Equilibrium" (with Drew Fudenberg and Giacomo Lanzani). 2024. *Journal of Political Economy*.

Dynamic Mechanism Design

Much of Philipp's early work was on dynamic mechanism design. His job market paper [3], written with Thomas Kruse, studies the optimal contract for a principal to give an agent who privately observes the value of a Markov chain (a stochastic process where the probability of the next event depends only on the current state) until the agent stops and obtains a reward that depends on the current level of the chain and on the stopping time. The principal can give the agent a transfer that depends only on the time the agent stops. The paper shows that in this setting, a stopping rule is implementable if and only if it takes the form of a cutoff rule, that is, "Stop if the current value is at least the cutoff value." Philipp has many other papers on other dynamic mechanism problems.

Philipp's work in these areas also includes [2],with Dirk Bergemann, which studies revenue maximization in a model where the preferences of consumers need not be fixed or even independent and identically distributed but can evolve stochastically. Gershkov, Moldovanu, and Strack [7] enrich the earlier Bergemann and Strack model to allow the mechanism designer to learn about the state of demand and for customers to strategically delay their purchases.

Majorization

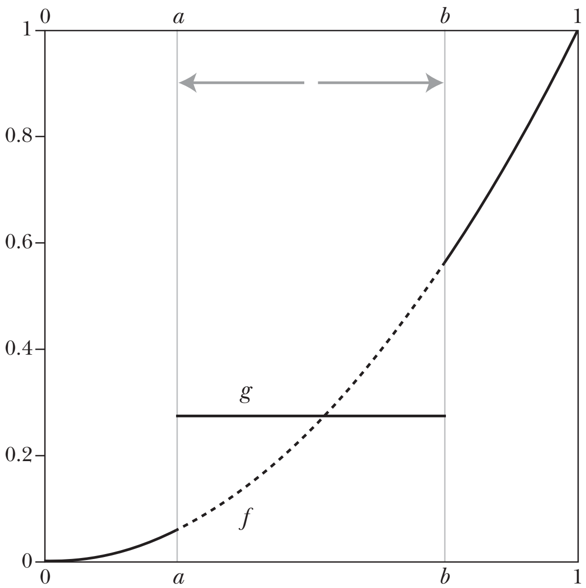
Kleiner, Moldovanu, and Strack [18] provide two powerful technical results that highlight a connection between many already-studied problems of mechanism and/or information design and can also be used to analyze new ones. The key observation is that many such problems reduce to selecting a best element from a constraint set that is given by a "majorization" condition. Hardy, Littlewood, and Polya (1929) said that a non-decreasing real-valued function f "majorizes" another non-decreasing real-valued function g if they both have the same domain and same mean and the integral of f from any point in the domain to its upper limit exceeds that of g. Intuitively, this means that f is more dispersed or "spread out" than g. In many economic applications, the objective function is convex, and standard arguments then show that the objective obtains its maximum at an extreme point—a point that is not a convex combination of other points in the set. Moreover, it is also often the case that feasible allocations are monotone—that is, they either consistently increase or consistently decrease—and that they are majorized by some particular allocation. Thus, in [18] the authors study the extreme points of sets of monotone functions that either majorize or are majorized by a given function. They show that any such extreme point is characterized by a collection of intervals on which the extreme point is piecewise constant, while outside the collection it coincides with the given function. This is illustrated in Figure 1, where the function f is majorized by the function g, *and* where the values of f points a and b have all been replaced by the average value of f over this interval.

One application of this approach is to feasible symmetric allocations in independent private value allocation problems, where all agents agree on the rankings of the objects, but some agents value every object more. The paper shows the *interim quantile allocation* of any feasible allocation is majorized by interim quantile allocation of the efficient rule "assign higher-ranked objects to types with higher expected valuations." (The interim quantile allocation is the map from the type's quantile to the expected quality it is allocated.) Then [18] notes that any allocation that is weakly majorized by the efficient allocation is dominant-strategy incentive compatible. Thus, all feasible and Bayesian incentive-compatible allocations can be implemented in dominant strategies, as in the case of identical objects studied by Gershkov et al. (2013). The paper next analyzes the maximization of linear functions given a majorization constraint and applies these results to show that the revenue-maximizing auctions for ranked objects can be implemented by a first-price auction with a reserve price when the distribution of types satisfies a regularity condition. The paper also applies its results to Bayesian persuasion, optimal delegation, and matching. This approach sheds new light on these widely studied applications while providing shorter and more revealing proofs.

Applications of Majorization

In work with Alex Gershkov, Benny Moldovanu, and Mengxi Zhang, Philipp used majorization and the extreme points approach to extend classical methods of mechanism design to frameworks that allow agents not to have expected utility preferences. In [22], these authors derive the optimal auction when bidders have non-expected utility preferences that exhibit constant risk aversion (that is, preferences between different risky options are the same regardless of the absolute scale of the potential outcomes), an assumption that allows for behavioral factors like

Figure 1
An Illustration of Majorization



Note: In this figure, the function *f* is majorized by the function *g*, and the values of *f* points *a* and *b* have all been replaced by the average value of *f* over this interval.

loss aversion (Köszegi and Rabin 2006), disappointment aversion (Gul 1991), and dual preferences (Yaari 1987). The paper shows that in the optimal mechanism, the utility of each agent is independent of the reports of agents, even though Maskin and Riley (1984) show that this is not generally the case with expected utility bidders who are strictly risk-averse.

In another application of majorization techniques, Gershkov, Moldovanu, Strack, and Zhang [24] solve the optimal insurance design problem for agents that face a loss of random magnitude and have "dual preferences" in the sense of Yaari (1987). In this setting, optimal contracts take menus of deductibles when agents have private information about the probability of losses and menus of coverage limits when the private information is about the magnitude of losses. Coverage limits on insurance are often observed in practice, yet they are not explained by the classic insurance model.

Yet another application focuses on auction design. Gershkov, Moldovanu, Strack, and Zhang [22] show that auctions where *c* bidders can make an investment after learning their type, but before bidding in the auction, can be viewed as auctions without the investment stage but instead a form of non-expected utility. This lets the authors apply majorization techniques to show that asymmetric equilibria may maximize the seller's expected revenue even when the bidders are ex-ante symmetric. The paper shows when symmetry is revenue-maximizing and that when it is, the optimal mechanism can be implemented as either a uniform-price auction or a discriminatory pay-your-bid auction with reserve prices that depend on the number of bidders and the number of units for sale.

In addition to these applications, the extreme-points approach has been used in newer work to study other issues: mechanism design with redistributive concerns (Akbarpour, Dworczak, and Kominers 2024), selling hard information (Ali et al. 2022), and nonlinear pricing by a seller seeking to maximize their worst-case profit (Bergemann, Heumann, and Morris 2023).

Information Economics

The study of asymmetric information and adverse selection has shed light on many economic phenomena and institutions since it took off in the 1960s. Much of this work takes the information the agents have as given, despite Arrow's (1996) complaint that "[n]ot enough weight has been given to the possibility that information can be . . . altered by economic decisions." Philipp's work has illuminated some important effects of the fact that much of peoples' information is the result of their own decisions, and helped rekindle research on this topic. He has also made important contributions to the study of how to reveal useful information while respecting privacy constraints.

Studying the choice of information structure requires specifying what sort of information can be obtained and the cost of obtaining it. The question of which cost functions are reasonable has been underexplored, and Philipp and coauthors have done important work on this as well.[1]

The Drift-Diffusion Model

A large literature in neuroscience and psychology uses the *drift-diffusion model* to explain the joint distribution of choices and decision times in laboratory experiments. (This model is a continuous-time version of the hypothesis-testing problem in Wald (1947), where an agent gathers costly samples before stopping and either accepting or rejecting the null hypothesis.) In the classic version of the drift-diffusion model, the agent observes a diffusion process with known volatility and uncertain drift until the agent takes one of two possible actions, where both the payoff and the drift of the signal are determined by the unknown state of the world, which has only two possible values. The drift-diffusion model was first proposed as a model of choice processes in perception tasks, where the subjects are asked to identify visual or auditory stimuli, for example, "Are more of the geese on this screen moving left or moving right?" The model has subsequently been applied to choice experiments, where subjects choose from a set of consumption goods presented to them; for examples, see Krajbich et al. (2012) and Reutskaja et al. (2011), where

[1] Arrow (1985) complained about this, and relatively little progress has been made until fairly recently.

participants were asked to fast before the experiment, presented with several binary choices between food items, and then allowed to eat the food item they had chosen in a randomly selected trial.

The assumption that there are only two states of the world makes it easy to find the optimal stopping rule and was natural for Wald's hypothesis-testing problem, but in the drift-diffusion model it leads to the conclusion that the accuracy of the agent's choices is uncorrelated with when they stop. This conflicts with the data from many laboratory experiments, which typically find that earlier decisions in the same task are more likely to be correct (for example, Churchland, Kiani, and Shadlen 2008; Ditterich 2006). To better understand this regularity, Philipp, Thomas Strzalecki and I propose in [8] a variant of the drift-diffusion model where the agent has normally distributed independent priors on the utility of each choice, and the drift of the signal depends on the difference in the utility across these choices. In this model, unlike the classic drift-diffusion model, signal strength depends on the utility difference or on the ease of the perceptual task. Here, unlike with only two states, beliefs update more slowly as data accumulate. Thus, if the agent has been observing the process for a long time and the posterior means of the two choices are close, there are two reasons to stop sampling that are not present in the binomial model: The agent does not expect to learn much more, and the two actions are probably about as good.

In [8], we first show that an arbitrary stopping rule in a drift-diffusion model with a fixed drift leads to a negative correlation between accuracy and stopping time if the value of the diffusion that makes the agent want to stop—that is, the "stopping boundary"—is decreasing in time. Intuitively, in this case, when the agent stops quickly, it is because they received a very strong signal. Moreover, accuracy can be decreasing in decision time even if the stopping boundary on each trial is not monotone decreasing when the data consist of trials with randomly drawn values of the drift, because averaging over trials creates an additional selection effect beyond the one that occurs when the value of the drift is fixed: In trials where the agent chose quickly, the drift was probably larger than usual, so the agent received sharper signals and is more likely to be correct.[2] Finally, the aggregate data have a speed-accuracy complementarity when the agent uses an optimal stopping rule and their prior over the states is correct.[3] The paper also considers the case where the agent has a fixed amount of attention and decides how much to focus on each signal. The optimum is to devote equal attention to both signals at every point in time, which reduces to the main model.

Many experimental studies of accuracy and decision time fit the data with arbitrary stopping boundaries, instead of boundaries that are derived with a theoretical foundation. In [14], Philipp, Whitney Newey, Strzalecki, and I characterize the choice probabilities that can be generated by any stopping boundary and then show

[2] This assumes that the boundary does not increase "too quickly."

[3] Alternative explanations for this complementarity include time-varying costs (Drugowitsch et al. 2012) and endogenous time variation in signal intensity (Woodford 2014).

how to estimate the drift and the boundary and construct a test statistic. Subsequent work on optimal stopping and stochastic choice includes Liang, Mu, and Syrgkanis (2022), Hébert and Woodford (2023), Gonçalves (2024), and Che and Mierendorff (2019).

The Comparison of Experiments

When will an agent prefer one source of information to the other if they both have the same cost? Blackwell (1951) and the subsequent literature modeled information sources as maps from states to signals, and called them "experiments." One experiment is called a "garbling" of another if it can be generated by adding noise to the other structure. Blackwell proves that experiment K is (weakly) preferred to experiment K' for all decision problems if and only if K' is a garbling of K. Roughly speaking, a garbling amounts to adding noise, and it is relatively easy to show that this noise cannot help the decision-maker; the deeper part of the result is that the two experiments can be pairwise ranked over *all* possible action spaces, priors, and utility functions only if one is a garbling of the other. When K' is a garbling of K, we say that K "Blackwell-dominates" K'. This dominance notion of has been very influential and has been used to gain insights into mechanism design (Grossman and Hart 1983), as well as accounting (Demski 1973), game theory (Kandori 1992), and "Bayesian persuasion" (Kamenica and Gentzkow 2011).

Blackwell's (1951) result is for a single experiment, but experiments are usually not done in isolation. Blackwell (1953) asked whether there are experiments that are not ranked by garbling but become ranked when many samples are taken from each, and Torgersen (1970) gave an example. In [19], Philipp, Xiaosheng Mu, Luciano Pomatto, and Omer Tamuz characterize when this occurs in the case of binary states. Their very clever proof uses the idea of the Rényi divergence, which is an easily computed parametric measure of "informativeness" that measures the amount of information gained if one distribution is substituted for another. The authors show that if K Blackwell-dominates K', it dominates K' in the Rényi order for any parameter t. Importantly, the Rényi t-divergence of n independent trials is n times the t-divergence of one trial, so if n draws from K Blackwell-dominates n draws from K', then K Rényi-dominates K'. The authors show that under mild technical conditions, the converse holds as well when n is sufficiently large, so if K Rényi-dominates K', every decision maker prefers a large sample from K to a large sample from K'. The proof starts from the fact that the posterior belief after a sequence of independent and identically distributed experiments is a function of the prior and the product of the likelihood ratios for each observation—or equivalently, the sum of the log-likelihoods. It then shows that two experiments are Blackwell-ordered if and only if one of the associated "perfected log-likelihood ratios" first-order stochastically dominates the other. To complete the proof, the paper uses a sharp large deviations bound to relate the Rényi order to the perfected log-likelihood ratios.

This beautiful result raised some interesting questions: When will n draws from K and m draws from L be preferred to n draws from K^1 and m draws from L^1? And what can be said about comparing large experiments when there are more than two states? The first of these is (as far as I know) still open; the second has recently been answered by Farooq et al. (2024).

Social Learning

Most of the social learning literature assumes either that each agent only acts once (as in the models of herding and information cascades in Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992), or that the agents use a boundedly rational heuristic (as in DeGroot 1974; Ellison and Fudenberg 1993). Philipp and coauthors have extended the analysis of social learning by fully rational agents who act and learn over multiple periods. In [21], with Matan Harel, Elchanan Mossel, and Omer Tamuz and in [28], with Wanying Huang and Tamuz, they study settings where agents receive exogenous signals that are independent and identically distributed so that each agent will eventually learn the true state, whether or not they observe the actions of others. The papers focus on the rate of social learning and show that the speed is bounded above, regardless of how many players there are, even though the rate of learning is proportional to the number of players when all signals are public. The underlying intuition is that if individuals learned quickly, they would quickly reduce the amount they update their beliefs based on their private signals so that at some point, little or no new social information could be gained.

In [21], the focus is on a model with two states, myopic agents, a continuum of signals, and a fully connected social network—that is, each agent observes the actions of all others. In [28], the authors broaden this setting to allow any finite number of states, non-myopic agents, a finite number of signals, and more general networks, though the main result is for the strongly connected case where there is an observational path between every pair of agents. As in most of the social learning literature, both papers assume that the agents' payoff functions do not depend on the actions of the other agents. Both papers also assume that there is a unique optimal action in each state and an upper bound on the informativeness of any signal.[4]

The first steps in [28] are to show that after some random time, all agents behave myopically and that in every strongly connected network, all agents learn at the same rate so the asymptotic learning rate is the same as with myopic agents and a fully connected network.[5] For each agent, their decision process involves their own experience of a given state and a "public" term that describes the posterior belief of an "outsider" with the same priors who observes all of the actions. It follows that the outside observer must learn at least as quickly as society does and that there is an upper bound on the maximum informativeness of the signal from any

[4] The model in [21] allows for a continuum of signals, while [28] assumes the set of signals is finite.

[5] To prove this main result, the paper writes the log-likelihood ratios for each player and pair of states as the sum of the player's log-likelihood of the states and a "public" term that sums the prior log-likelihood and the history of actions the player has observed, where the public term describes the posterior log-likelihood of an observer who has the same prior and observes the same actions.

agent. The main theorem then follows from the fact that if the equilibrium speed of learning were higher than this maximal private rate, an outside observer would know what action an agent would take before seeing it—and so not learn anything more from that agent's actions. This would imply that the agents also stop learning from their private signals, but this is a contradiction because such learning is needed for the equilibrium to emerge at all. This finding is reminiscent of the Grossman and Stiglitz (1980) argument that a market cannot be efficient unless agents have incentives to collect information and to act on it, but in an efficient market, agents do not have an incentive to collect information because they cannot improve on the market outcome. (Interested readers should know that the proofs in this paper are particularly short and clear, perhaps because they proceed from first principles: the "heaviest" result used is that the learning rate for a single agent is well-defined.)

Privacy Constraints

Some settings have legal or ethical constraints on revealing personal information. For example, in the case of bank loans, the Equal Credit Opportunity Act "prohibits creditors from discriminating against credit applicants on the basis of race, color, religion, national origin, sex, marital status, age . . ." It is straightforward to prevent decisions from explicitly conditioning on one factor, like race, but other information such as zip code may be correlated with that factor. For this reason, the "algorithmic fairness" literature studies a more stringent requirement: the information the bank uses to make loan decisions must be statistically independent of an individual's protected characteristics.

In [30], Strack and Kai Hao Yang characterize the experiments that are privacy-preserving with respect to a collection of "privacy sets," meaning that after every realization of the experiment's signal, the posterior probability of each privacy set is the same as its prior. In the bank loan example, if the privacy sets correspond to the applicant's race, the posterior probability that the agent is a given race must be the same as the prior whether or not the signal says to grant the loan, which is perhaps too demanding a requirement. For this reason, Strack and Yang extend their analysis to signals that are *conditionally privacy preserving* signals, which only requires that the posterior is the same as the prior conditional on some "materially relevant information" that may be correlated with race (for example, default history).

In general, privacy sets need not be defined as a collection of states, but in [30], they show it is without loss of generality to assume that they are, which greatly simplifies the analysis. The paper shows that when the state is one-dimensional, a signal is privacy-preserving if and only if it is Blackwell-dominated by some "reordered quantile" signal. It shows that these reordered quantile signals are Blackwell-undominated among privacy-preserving signals. It then proves a version of the same result for general state spaces, for example, cases where the state includes both a measure of ability or fitness for a position and information about protected characteristics, which is an important generalization of the two-state case analyzed in He, Sandomirskiy, and Tamuz (2022). Finally, the paper shows how to find the best privacy-preserving (or conditional privacy-preserving) signal for any objective function, and finds a closed-form solution when the objective has "increasing differences" (for example, if the objective is to minimize the expected distance between the action and the state).

The Cost of Information

What sorts of information cost functions are economically reasonable? Early work on costly information acquisition assumed a parametric cost function for signals that are increasing in a parameter that measures the informativeness or accuracy of a signal, but it can be hard to tell which particular specification best fits a given setting. This has led to an interest in more abstract cost functions that do not specify a particular set of signals, such as "entropy costs," which measure the expected divergence between the agent's posterior beliefs and their prior (for comparison, see Arrow 1985; Sims 2003). In [26], Pomato, Strack, and Tamuz provide an axiomatic characterization of information cost functions in a setting with a finite number of possible states of the world where the likelihood ratios for each pair of states are finite almost surely. It shows that a cost function that satisfies four axioms must be a *LLR cost*, which means that it is a weighted sum of the expected log-likelihood ratios between pairs of states, where the weight on the log-likelihood between state i and state j can depend on the prior probability of state i.

They build on four axioms. The first axiom requires that two experiments that are Blackwell-equivalent have the same cost. This axiom will be automatically satisfied when experiments are chosen optimally, as all decision-makers prefer the Blackwell-equivalent experiment with the lowest cost. Second, the cost of running two independent experiments is the sum of their costs, which implies that the time taken to gather information has no cost beyond the number of samples. Third, if an experiment either succeeds (and produces an informative signal) or fails and yields no information, the marginal cost of increasing the probability of success is constant. This axiom will be satisfied if the information-seeking agent can create "diluted" experiments that randomize between a fixed signal structure and an uninformative signal, and the cost of an experiment is not greater than the expected cost of performing its diluted version until an informative signal is generated. It is satisfied by a wide range of possible cost structures. Fourth, the cost function is continuous in the total variation norm, which measures the absolute value of upward or downward movements in the underlying function.

To illustrate the differences between LLR cost functions and entropy, consider an agent who sees red and blue dots on a video screen and is uncertain whether there are x red dots and 100-x blue dots or 100-x red dots and x blue dots. Here the entropy cost of a symmetric binary signal that either says "there are more red dots" or "there are more blue dots" is independent of x, while the LLR cost need not be. Next, suppose the agent thinks both states are equally likely, and consider the two costs as functions of the binary signal's precision—that is, the probability that it reports the correct state. Then both costs are 0 for the completely uninformative signal that reports "more red" with probability 1/2, regardless of the state. However, as the probability the signal reports the true state goes to 1, the LLR cost goes to infinity (because the value placed on the prior goes toward 0, so the ratio increases), while the entropy cost converges to ln(2): LLR's use of log-likelihood suggests a much higher cost for very accurate messages.

Imperfect Learning and Behavioral Economics

A standard approach in behavioral economics is to modify one or two assumptions of a standard theory in the direction of greater psychological realism (Camerer, Loewenstein, and Rabin 2004). However, as I noted in Fudenberg (2006), the evidence that supports one modification from the standard model may also argue for additional modifications. Moreover, the great multiplicity of behavioral theories can make it difficult to know which ones to apply in a given situation or how they interact. As I argued, both of these problems could be mitigated if economists could find a way to integrate the many possible assumptions of behavioral economics into a few more general models. Philipp's work on imperfect learning has provided a major step in this direction. This work has had two main focuses: learning by agents who have perfect memory but whose prior beliefs prevent them from learning the truth, and agents whose main difficulties in learning come from their imperfect memory. In both cases, Philipp develops models that link learning directly with belief and behavior and generate realistic and interesting behavioral biases.

Misspecification and Learning

Economists typically assume that agents are correctly specified, meaning that their prior beliefs do not rule out the truth. Such agents will eventually learn the consequences of any action they play infinitely often, although they need not learn the consequences of actions they rarely or never play. Yet people sometimes seem to have incorrect beliefs despite having an abundance of relevant data, such as a belief that taxes are linear in income when they are not, or "causation neglect" about the impact of actions on outcomes. One explanation for this is that they may be *misspecified*, and have prior beliefs that rule out the true state of the world. Such misspecification is especially relevant for parametric learning models, as any parametric prior (such as the assumption of a linear demand curve with unknown slope and intercept) assigns a probability of zero to "most" data-generating processes.

To understand the effect of misspecified priors, recall that under Bayesian updating, the posterior odds ratio between states is their prior odds ratio multiplied by the likelihood ratio the agent assigns to the empirical distribution. Thus, when agents are correctly specified and their data identify the true data-generating process, their beliefs converge to the truth, as it is the only model that maximizes the empirical likelihood. Moreover, a similar argument shows that the beliefs of a misspecified agent asymptotically concentrate on the set of models that maximize the likelihood of the observed data when the data-generating process is exogenous and sufficient to identify the true data-generating process.

However, in many economic applications, actions and associated signal distributions depend on the action taken by the agent, and knowing the probability distribution over outcomes that one action generates may not pin down the probability distributions generated by other actions. This interdependence between the agent's misspecification, beliefs, and long-run actions makes the dynamics of misspecified learning much richer. In an early analysis of this issue, Arrow and Green (1973) develop a general model of learning by misspecified agents who always play a myopic best response to their current beliefs. This type of "passive learning" is optimal if agents do not care about their future payoffs, or if the information they observe does not depend on their actions. Esponda and Pouzo (2016) defined Berk-Nash equilibrium, in which agents play a best response to (possibly mistaken) beliefs that best match the data generated by the true data-generating process. They relate this new equilibrium concept to the asymptotic behavior of passive learning by misspecified agents by combining Berk's (1966) work on misspecified learning with the Fudenberg and Kreps (1993) analysis of learning in games. The formulation of Berk-Nash equilibrium led to a resurgence of interest in misspecified learning and helped to inspire several of Philipp's papers.

Learning by Overconfident Agents

In [9], Strack, Paul Heidhues, and Botond Köszegi study passive learning by an agent whose misspecification takes the form of a dogmatic belief that their ability is higher than it really is. The agent's output depends on their ability, their effort, and an unknown "fundamental" that corresponds to the objective productivity of the task. Given their observed relatively low performance, the agent's overconfidence in their own ability leads them to believe that the fundamental is lower than they had thought, and thus to reduce their effort. Over time the agent's beliefs become monotonically further from the truth, and their action gets farther from optimal. As the paper notes, this provides another illustration of how assuming that the overconfident agent is otherwise "classical" can be misleading.[6]

In a twist on this setting in [17], the same three authors modify the observation structure of their earlier paper in [9], so that the agent's inferences do not depend on their actions, which ensures that passive learning is optimal even if the agent is not myopic. The paper then uses stochastic approximation to show that beliefs and actions converge. Unlike the earlier model, this version allows for multiple steady states, so even if the system almost surely converges, what it converges to may be stochastic.

[6] To prove that beliefs and actions converge, Heidhues, Köszegi, and Strack develop in [9] a version of the martingale law of large numbers for Markov chains that replaces the usual square-integrability condition with a restriction to real-valued chains that grow sub-linearly almost surely, a result that will be useful more generally.

Excessive Experimentation

In [5], Philipp, Gleb Romanyuk, and I published the first study of the dynamics of learning by a misspecified agent who is patient and thus is willing to engage in "active learning" by experimenting with actions that do not maximize their expected current payoff. If an expected-utility maximizing agent is correctly specified, once they have lots of data, they do not expect any other action will provide much new information, so they play the action that maximizes their current period's payoff, and their actions eventually converge. In contrast, we show in [5] that the agent's actions can cycle instead of converge when they care about future payoffs and are misspecified, even if they would converge if the agent is myopic. This novel interaction between patience and misspecification arise because misspecification leads the agent to continue to believe they could learn a lot from continued experimentation with other actions.

Convergence

In [20], Philipp, Giacomo Lanzani, and I provide a sharp general characterization of the "limit outcomes" of misspecified learning by agents who need not be myopic. The paper first shows that any limit point must be a uniform Berk-Nash equilibrium, where "uniform" means that the limit action must be the best response to any mixture of sample realizations that maximize the empirical likelihood over the models in the support of the agent's prior. To show this, the paper first extends past work on the rate that beliefs converge to allow for misspecification. It then shows that if play converges to a particular action, the empirical outcome frequencies converge to the corresponding data-generating process with oscillations that die out more slowly than the rate at which beliefs concentrate about the data. Thus, if the limit action were not a best reply to one of the likelihood-maximizing models, the agent would switch to another action when the beliefs concentrate on that maximizer. Finally, the paper shows that an action has a high probability of being the limit outcome for "nearby" beliefs if and only if it is a "uniformly strict Berk-Nash equilibrium," meaning that is a strict best reply to every mixture over the relevant maximizers. The key step here is showing that if beliefs initially assign a very high probability to models that make the limit action optimal, they are unlikely to drop below the threshold that makes it suboptimal.[7]

Say that an equilibrium is "positively attractive" if it has a positive probability of being the long-run outcome from any starting belief. Uniformly strict Berk-Nash equilibria are positively attractive under various types of misspecification, including various forms of incorrect perceptions of how actions influence the distribution of outcomes, such as a belief that outcomes are exogenous when they are not, or that outcome distribution under one action is uncorrelated with that under another. The paper uses this framework to show that there can be a positive probability that

[7] The proof generalizes an "upcrossing argument" of Fudenberg and Levine (1992) to misspecified beliefs and then strengthens a bound from Frick, Iijima, and Ishii (2020) to hold uniformly.

a patient but misspecified agent will fall into an "underinvestment trap" when a correctly specified agent would not.

Selective Memory

For most people, memory is limited and selective: They do not remember everything and are more likely to remember some things than others (for example, Shadlen and Shohamy 2016; Schacter 2008; Kahana 2012). In [31], Philipp, Giacomo Lanzani, and I study the implications of selective memory when a myopic agent chooses actions that maximize expected utility. We assume that agents compute their beliefs from their prior and remembered experiences, as opposed to a full update of their posterior beliefs, and that agents are unaware of their selective memory. Thus, the agents update their beliefs as if the experiences they remember are the only ones that occurred.[8] The paper also assumes that agents choose actions that maximize the expected payoff in the current period. We show that if the agent eventually always plays a given action, then that action can be viewed as a *selective memory equilibrium*, meaning that the action is a best response to a belief that maximizes the likelihood of a version of the outcome distribution that gives more weight to the outcomes the agent is more likely to remember. Using stochastic approximation results from Benaïm, Hofbauer, and Sorin (2005), Esponda, Pouzo, and Yamamoto (2021), and a result on the rate that beliefs converge from our earlier paper [25], we show in [31] that every limit strategy is a selective memory equilibrium. It also provides a sufficient condition for beliefs and actions to converge.

The effects of selective memory can resemble the effects of misspecification. Indeed, [31] shows that most long-run outcomes that can be achieved as the result of misspecified learning can also be induced by selective memory and vice-versa (with some subtleties involved when the equilibria in question are not uniformly strict). Importantly, the form of misspecification that would lead to the same behavior as a given form of selective memory depends on the environment. That is, particular forms of misspecification and selective memory that coincide under one information structure can respond very differently to changes in what the agent observes. For example, combining positive and negative feedback has different effects on agents with ego-boosting memory than on agents who are mistakenly sure that they have high ability. The paper shows that "associativeness" has no effect on long-run outcomes and that ego-boosting memory bias leads agents to underestimate the ability of their coworkers.

[8] For evidence that agents access their accumulated evidence each period when updating belief, see d'Acremont, Schultz, and Bossaerts (2013) and Sial, Sydnor, and Taubinsky (2023). In addition, Reder (2014), Zimmermann (2020), and Gödker, Jiao, and Smeets (2022) provide evidence of partial or complete unawareness of memory biases.

Risk Preferences

Risk preferences are a key factor in the insurance and investment markets and also matter for decisions related to health, job choice, and retirement. For expected utility agents, risk preferences depend on the totality of risks they face—as opposed to the properties of each risk in isolation—so it is important to understand how risk preferences change as additional risks are added. It is also important to understand the behavior of agents who do not maximize expected utility, but instead evaluate risks using nonlinear functions of their probabilities. Philipp has done impressive work on both topics.

Adding Noise

The concepts of first- and second-order stochastic dominance play a key role in the economics of risk and insurance. To think about these concepts, compare two distributions or "lotteries" p and q over real-valued outcomes. Lottery p first-order stochastically dominates lottery q if p is more likely than q to be above any fixed value z, or equivalently when the expectation of f with respect to p is larger than its expectation with respect to q for every increasing function f for which the expectations are well defined. Not all pairs of lotteries can be ranked in this way, but when they can, every expected utility agent who prefers more money to less prefers the stochastically-dominant one. Similarly, lottery p second-order stochastically dominates lottery q if and only if it is weakly preferred by all agents with weakly increasing and weakly concave preferences. Here the concave utility function models risk aversion: It allows agents to prefer lotteries with lower expected value but less risk. [9]

Adding noise to lotteries preserves stochastic dominance: If lottery p first- or second-order stochastically dominates lottery q, the same is true for the compound lotteries $p + r$ and $q + r$ for any lottery r that is independent of p and q. However, the converse is not true: Even if p does not first- or second-order dominate q, there can be an independent lottery r such that $p + r$ does dominate $q + r$.

More strongly, Pomatto, Strack, and Tamuz [13] show that if the expected value of a random variable x is higher than that of y, there is an independent random variable r such that $x + r$ first-order stochastically dominates $y + r$. Also, if lotteries p and q have the same mean and the variance of p is strictly less than that of q, then there is an independent lottery r such that $p + r$ second-order stochastically dominates $q + r$.[10] It is important here that r can take on arbitrarily large and small values, but they argue that the required standard deviations need not be implausibly large. For example, if p pays 12 and -10 with probability $1/2$ each, then it does not first-order

[9] When p and q in addition have the same mean, the second lottery q is a mean-preserving spread of p (Rothschild and Stiglitz 1978) and an example of the majorization relation discussed above.

[10] Tarsney (2018) independently found an example of this for first-order dominance. In [29], Mu, Pomatto, Strack, and Tamuz generalize the earlier paper [13] by Pomatto, Strack, and Tamuz to other statistics than moments, and provide a sharper result for the case where the added random variable is required to be bounded.

stochastically dominate 0, but there is an r with large standard deviation (3,525) such that $p + r$ first-order stochastically dominates $0 + r$.

The idea of the proof for first-order dominance is as follows. The assumption that the expected value of x is higher than that of y implies that the cumulative distribution function of y is on average above that of x. This suggests that adding a lottery r with a "sufficiently diffuse" distribution will lead to the cumulative distribution function of $y + r$ being above that of $x + r$ at least at most points. To show there is a lottery r that makes the cumulative distribution function of $y + r$ everywhere above that of $x + r$ they make clever use of a result of Ruzsa and Szekely (1988), which shows that a signed measure that assigns total mass 1 to R can be "smoothed" into a probability measure by "convolving" it (that is, combining the two functions) with an appropriately chosen probability measure.

Background risk refers to risk that the agent will face whatever choices they make. An immediate corollary of the proof in [13] of first-order dominance is that for each finite set of gambles, there is a background risk such that any agent whose preferences respect first-order stochastic dominance will rank gambles by their expected value. A related result in [27] from Mu, Pomatto, Strack, and Tamuz is that an agent who faces unbounded background risk will accept any gamble that has positive expectation and whose riskiness in the sense of Aumann and Serrano (2008) is less than a measure of the "size" of the background risk. It uses this finding to argue that under plausible levels of background risk, expected utility theory, prospect theory, and rank-dependent utility all imply that agents who account for background risk and respect stochastic dominance are risk-neutral over small gambles.

Cumulative Prospect Theory in a Dynamic Setting

Cumulative prospect theory (as described in Tversky and Kahneman 1992) is the most widely used generalization of expected utility theory. It supposes that outcomes are evaluated relative to a "reference point" that separates outcomes into gains and losses, and the agent distorts cumulative probabilities so that various sizes of gains and losses receive disproportionate weight in the utility function, as in Quiggin (1982). Most applications of cumulative prospect theory further assume that the agent overweights extremely unlikely gains and underweights extremely likely losses. Under fairly general conditions, an agent who distorts probabilities in this way will accept some small binomial gambles with negative expected payoff, provided their loss aversion is not too strong.[11]

In [3], Philipp and Sebastian Ebert use this fact about static choice to study an agent who needs to decide when to stop a diffusion process with a known negative drift and claim the current state as their payoff. A risk-averse expected utility maximizer would stop immediately (because of the known negative drift), but they show that a "naïve" agent with cumulative prospect theory preferences will never

[11] Azevedo and Gottlieb (2012) show this is true for sufficiently large gambles.

stop. Here "naïve" means that the agent chooses their stopping rule believing it will be carried out, when in fact at subsequent periods they will choose to continue; that is, the agent is dynamically inconsistent in the sense of Strzalecki (2024). For such an agent, a simple stopping rule with asymmetric upper and lower boundaries can generate negative-expected-value binary lotteries that the agent finds appealing. Conversely, the same two authors show in [6] that a "sophisticated" cumulative prospect theory agent who is aware that their future selves will have different preferences cannot commit to future behavior, and for this reason may never start a diffusion process with high positive drift because they are afraid that the future selves will stop too soon. It concludes that in dynamic settings, prospect theory needs to be modified to yield more realistic predictions.

Conclusion

Philipp is one of the most cheerful and friendly people I know. I have greatly enjoyed working with and learning from him. He is also extraordinarily productive: as of June 2024, he had published 37 papers in the eleven years since the completion of his dissertation, with several others forthcoming. Of this output, my review has only covered some of the highlights. In addition to the work discussed here, Philipp has made contributions to the study of market design (in [11], with Jacob Leshno and [23], with Afshin Nikzad), on contests (in [15], with Dawei Fang and Thomas Noe and [1], with Christian Seel), on present bias (in [10], with Paul Heidhues), and on epidemics (in [12], with Thomas Kruse).

Some of Philipp's papers provide a new understanding of important economics phenomena, others introduce results and techniques that will be used for years to come, and many do both. All of his work is clear and insightful, and its influence seems sure to expand.

■ *I thank Roberto Corrao, Yifan Dai, David Kreps, Annie Liang, Giacomo Lanzani, Benny Moldavanu, Xiaosheng Mu, Luciano Pomato, Larry Samuelson, Philipp Strack, and Omer Tamuz for helpful comments, and National Science Foundation grant SES-24171632 for financial support.*

References

Akbarpour, Mohammad, Piotr Dworczak, and Scott Duke Kominers. 2024. "Redistributive Allocation Mechanisms." *Journal of Political Economy* 132 (6): 1831–75.

Ali, S. Nageeb, Nima Haghpanah, Xiao Lin, and Ron Siegel. 2022. "How to Sell Hard Information." *Quarterly Journal of Economics* 137 (1): 619–78.
Arrow, Kenneth J. 1985. "Informational Structure of the Firm." *American Economic Review* 75 (2): 303–07.
Arrow, Kenneth J. 1996. "The Economics of Information: An Exposition." *Empirica* 23: 119–28.
Arrow, Kenneth J., D. Blackwell, and M. A. Girshick. 1949. "Bayes and Minimax Solutions of Sequential Decision Problems." *Econometrica* 17 (3/4): 213–44.
Arrow, Kenneth J., and Jerry R. Green. 1973. "Notes on Expectations Equilibria in Bayesian Settings." Stanford University Working Paper 33.
Aumann, Robert J., and Roberto Serrano. 2008. "An Economic Index of Riskiness." *Journal of Political Economy* 116 (5): 810–36.
Azevedo, Eduardo M., and Daniel Gottlieb. 2012. "Risk-Neutral Firms Can Extract Unbounded Profits from Consumers with Prospect Theory Preferences." *Journal of Economic Theory* 147 (3): 1291–99.
Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107 (3): 797–817.
Benaïm, Michel, Josef Hofbauer, and Sylvain Sorin. 2005. "Stochastic Approximations and Differential Inclusions." *SIAM Journal on Control and Optimization* 44 (1): 328–48.
Bergemann, Dirk, Tibor Heumann, and Stephen Morris. 2023. "Cost Based Nonlinear Pricing." Cowles Foundation Discussion Paper 2368.
Berk, Robert H. 1966. "Limiting Behavior of Posterior Distributions When the Model Is Incorrect." *Annals of Mathematical Statistics* 37 (1): 51–58.
Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026.
Blackwell, David. 1951. "Comparison of Experiments." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* edited by Jerzy Neyman, 93–102. University of California Press.
Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *Annals of Mathematical Statistics* 24 (2): 265–72.
Bohren, J. Aislinn, and Daniel N. Hauser. 2021. "Learning with Heterogeneous Misspecified Models: Characterization and Robustness." *Econometrica* 89 (6): 3025–77.
Border, Kim C. 1991. "Implementation of Reduced Form Auctions: A Geometric Approach." *Econometrica* 59 (4): 1175–87.
Camerer, Colin F., George Loewenstein, and Matthew Rabin. 2004. *Advances in Behavioral Economics.* Princeton University Press.
Caplin, Andrew, and Mark Dean. 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." NBER Working Paper 19318.
Che, Yeon-Koo, and Konrad Mierendorff. 2019. "Optimal Dynamic Allocation of Attention." *American Economic Review* 109 (8): 2993–3029.
Churchland, Anne K., Roozbeh Kiani, and Michael N. Shadlen. 2008. "Decision Making with Multiple Alternatives." *Nature Neuroscience* 11: 693–702.
d'Acremont, Mathieu, Wolfram Schultz, and Peter Bossaerts. 2013. "The Human Brain Encodes Event Frequencies While Forming Subjective Beliefs." *Journal of Neuroscience* 33 (26): 10887–97.
Degroot, Morris H. 1974. "Reaching a Consensus." *Journal of the American Statistical Association* 69 (345): 118–21.
Demski, Joel S. 1973. "The General Impossibility of Normative Accounting Standards." *Accounting Review* 48 (4): 718–23.
Denti, Tommaso. 2022. "Posterior Separable Cost of Information." *American Economic Review* 112 (10): 3215–59.
Ditterich, Jochen. 2006. "Stochastic Models of Decisions about Motion Direction: Behavior and Physiology." *Neural Networks* 19 (8): 981–1012.
Drugowitsch, Jan, Rubén Moreno-Bote, Anne K. Churchland, Michael N. Shadlen, and Alexandre Pouget. 2012. "The Cost of Accumulating Evidence in Perceptual Decision Making." *Journal of Neuroscience* 32 (11): 3612–28.
Ellison, Glenn, and Drew Fudenberg. 1993. "Rules of Thumb for Social Learning." *Journal of Political Economy* 101 (4): 612–43.
Esponda, Ignacio and Demian Pouzo. 2016. "Berk–Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models." *Econometrica* 84 (3): 1093–1130.
Esponda, Ignacio, Demian Pouzo, and Yuichi Yamamoto. 2021. "Asymptotic Behavior of Bayesian

Learners with Misspecified Models." *Journal of Economic Theory* 195: 105260.

Farooq, Muhammad Usman, Tobias Fritz, Erkka Haapsalo, and Marco Tomamichel. 2024. "Matrix Majorization in Large Samples." *IEEE Transactions on Information Theory* 70 (5): 3118–44.

Frick, Mira, Ryota Iijima, and Yuhta Ishii. 2020. "Stability and Robustness in Misspecified Learning Models." Cowles Foundation Discussion Paper 2235.

Frick, Mira, Ryota Iijima, Yuhta Ishii. 2023. "Belief Convergence under Misspecified Learning: A Martingale Approach." *Review of Economic Studies* 90 (2): 781–814.

Fudenberg, Drew. 2006. "Advancing Beyond *Advances in Behavioral Economics*." *Journal of Economic Literature* 44 (3): 694–711.

Fudenberg, Drew, Kevin He, and Lorens A. Imhof. 2017. "Bayesian Posteriors for Arbitrarily Rare Events." *Proceedings of the National Academy of Sciences* 114 (19): 4925–29.

Fudenberg, Drew, and David M. Kreps. 1993. "Learning Mixed Equilibria." *Games and Economic Behavior* 5 (30: 320–67.

Fudenberg, Drew, and David K. Levine. 1992. "Maintaining a Reputation When Strategies Are Imperfectly Observed." *Review of Economic Studies* 59 (3): 561–79.

Gershkov, Alex, Jacob K. Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. 2013. "On the Equivalence of Bayesian and Dominant Strategy Implementation." *Econometrica* 81 (1): 197–220.

Gödker, Katrin, Peiran Jiao, and Paul Smeets. 2022. "Investor Memory." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.3348315.

Gonçalves, Duarte. 2024. "Speed, Accuracy, and Complexity." Preprint, arXiv. https://doi.org/10.48550/arXiv.2403.11240.

Grossman, Sanford J., and Oliver D. Hart. 1983. "An Analysis of the Principal-Agent Problem." *Econometrica* 51 (1): 7–45.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70 (3): 393–408.

Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica* 59 (3): 667–86.

Hardy, G. H., J. E. Littlewood, and G. Polya. 1929. "Some Simple Inequalities Satisfied by Convex Functions." *Messenger Math* 58: 145–52.

He, Kevin, Fedor Sandomirskiy, and Omer Tamuz. 2022. "Private Private Information." Preprint, arXiv. https://doi.org/10.48550/arXiv.2112.14356.

Hébert, Benjamin, and Michael Woodford. 2023. "Rational Inattention When Decisions Take Time." *Journal of Economic Theory* 208: 105612.

Kahana, Michael Jacob. 2012. *Foundations of Human Memory*. Oxford University Press.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590–2615.

Kandori, Michihiro. 1992. "Social Norms and Community Enforcement." *Review of Economic Studies* 59 (1): 63–80.

Köszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics* 121 (4): 1133–65.

Krajbich, Ian, Carrie Armel, and Antonio Rangel. 2010. "Visual Fixations and the Computation and Comparison of Value in Simple Choice." *Nature Neuroscience* 13: 1292–98.

Krajbich, Ian, Dingchao Lu, Colin Camerer, and Antonio Rangel. 2012. "The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions." *Frontiers in Psychology* 3: 193.

Krajbich, Ian, and Antonio Rangel. 2011. "Multialternative Drift-Diffusion Model Predicts the Relationship Between Visual Fixations and Choice in Value-Based Decisions." *Proceedings of the National Academy of Sciences* 108 (33): 13852–57.

Liang, Annie, Xiaosheng Mu, and Vasilis Syrgkanis. 2022. "Dynamically Aggregating Diverse Information." *Econometrica* 90 (1): 47–80.

Maskin, Eric, and John Riley. 1984. "Optimal Auctions with Risk Averse Buyers." *Econometrica* 52 (6): 1473–1518.

Milosavljevic, Milica, Jonathan Malmaud, Alexander Huth, Christof Koch, and Antonio Rangel. 2010. "The Drift Diffusion Model Can Account for the Accuracy and Reaction Time of Value-Based Choices under High and Low Time Pressure." *Judgment and Decision Making* 5 (6): 437–49.

Myerson, Roger B. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1): 58–73.

Quiggin, John. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior and Organization* 3 (4): 323–43.

Reder, Lynne M., ed. 2014. *Implicit Memory and Metacognition*. Psychology Press.

Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel. 2011. "Search Dynamics in Consumer Choice under Time Pressure: An Eye Tracking Study." *American Economic Review* 101 (2): 900–926.

Roe, Robert M., Jermone R. Busemeyer, and James T. Townsend. 2001. "Multialternative Decision Field Theory: A Dynamic Connections Model of Decision Making." *Psychological Review* 108 (2): 370–92.

Rothschild, Michael, and Joseph E. Stiglitz. 1978. "Increasing Risk: I. A Definition." In *Uncertainty in Economics*, 99–121. Academic Press.

Ruzsa, Imre Z., and Gabor J. Szekely. 1988. *Algebraic Probability Theory. Wiley Series in Probability and Statistics*. Wiley.

Schacter, Daniel L. 2008. *Searching for Memory: The Brain, the Mind, and the Past*. Basic Books.

Shadlen, Michael N., and Daphna Shohamy. 2016. "Decision Making and Sequential Sampling from Memory." *Neuron* 90 (5): P927–39.

Sial, Afras Y., Justin R. Sydnor, and Dmitry Taubinsky. 2023. "Biased Memory and Perceptions of Self-Control." NBER Working Paper 30825.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.

Strassen, Volker. 1965. "The Existence of Probability Measures with Given Marginals." *Annals of Mathematical Statistics* 36 (2): 423–39.

Strzalecki, Tomasz. 2024. *Decision Theory*. http://datalaundering.com/download/notes.pdf.

Tarsney, Christian. 2018. "Exceeding Expectations: Stochastic Dominance as a General Decision Theory." Preprint, arXiv. https://doi.org/10.48550/arXiv.1807.10895.

Torgersen, Erik Nikolai. 1970. "Comparison of Experiments When the Parameter Space Is Finite." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 16: 219–49.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5: 297–323.

Wald, Abraham. 1947. "Foundations of a General Theory of Sequential Decision Functions." *Econometrica* 15 (4): 279–313.

Woodford, Michael. 2014. "An Optimizing Neuroeconomic Model of Discrete Choice." NBER Working Paper 19897.

Yaari, Menahem E. 1987. "The Dual Theory of Choice under Risk." *Econometrica* 55 (1): 95–115.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–61.

Journal of Economic Perspectives—Volume 39, Number 1—Winter 2025—Pages 247–254

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at <taylort@macalester.edu>, or c/o Journal of Economic Perspectives, Macalester College, 1600 Grand Ave., Saint Paul, MN 55105.

Smorgasbord

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2024 has been awarded to Daron Acemoglu, Simon Johnson, and James Robinson "for studies of how institutions are formed and affect prosperity." Each year, the Nobel Committee publishes a "Popular Information" overview of the award and a "Scientific Background" essay that goes into greater depth (at https://www.nobelprize.org/prizes/economic-sciences/2024). Here's one part of the explanation: "It is instructive to put the contribution of Acemoglu and Robinson in perspective and relate it to the literature that already existed in the late 1990s. . . . Recall that the standard answer to why elites gave up the control of economic and political institutions was embodied in modernization theory and related explanations. According to these theories, the process of socioeconomic development would eventually bring about democratization, essentially as a by product of economic progress. As societies become richer, this wealth brings about rising education, a more plentiful middle class, and gradually milder conflict over income inequality, factors which all favor democratization. A

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20241437.

second approach, which challenged modernization (and other structural) theories, argued that democratization is instead the by-product of patterns of strategic interaction among political elites. Personal skills, luck, or strategic mistakes are, according to this approach, part and parcel of what democratization is about. . . . While the second view thus holds that democracy is usually granted or undermined from above, a third approach to explaining democratization, by contrast, points to the importance of social forces in society, most importantly different class actors. The key assertion in this tradition is that democracy is imposed from below by the people through popular mobilization. According to this view, incumbent authoritarian elites would not care to enact reforms or bargain with the democratic opposition if they did not fear the masses or an imminent threat of revolution. Acemoglu and Robinson integrated these three traditions by providing structural conditions (such as economic crises), relating these to preferences over institutions and social forces (such as the threat of revolution), and by providing the conditions under which strategic elites chose to reform (such as extending the electoral franchise). This is one of the reasons why their approach has become so influential."

The *World Development Report 2024*, one of the flagship publications of the World Bank, discusses "The Middle Income Trap" (https://www.worldbank.org/en/publication/wdr2024). "Developing economies change in structure as they increase in size, which means that changes in the pace of growth stem from factors that are new to them. Although these imperatives can vary across countries, economic expansion, on average, begins to decelerate and often reaches a plateau in income per capita growth, typically at about 11 percent of US GDP per capita. Today, this figure would be about US$8,000, or around the level at which countries are firmly considered upper-middle- income. A systematic slowdown in growth then occurs. . . . [T]he pace of progress in middle-income countries is slowing. Average annual income growth in these countries slipped by nearly one-third in the first two decades of this century—from 5 percent in the 2000s to 3.5 percent in the 2010s. A turnaround is not likely soon because middle-income countries are facing ever-stronger headwinds. They are contending with rising geopolitical tensions and protectionism that can slow the diffusion of knowledge to middle-income countries, difficulties in servicing debt obligations, and the additional economic and financial costs of climate change and climate action. . . . To achieve more sophisticated economies, middle-income countries need two successive transitions, not one. In the first, investment is complemented with infusion, so that countries (primarily lower-middle-income countries) focus on imitating and diffusing modern technologies. In the second, innovation is added to the investment and infusion mix, so that countries (primarily upper-middle-income countries) focus on building domestic capabilities to add value to global technologies, ultimately becoming innovators themselves. In general, middle-income countries need to recalibrate the mix of the three drivers of economic growth—investment, infusion, and innovation—as they move through middle-income status."

John A. List explores concerns over "Field Experiments: Here Today Gone Tomorrow?" (*American Economist*, 69:2, published online August 6, 2024, https://journals.sagepub.com/doi/abs/10.1177/05694345241261340). He discusses difficulties of moving from simple field experiments to scaled-up policy advice. "In the A/B experimental test of an early childhood program summarized . . . the program is found to triple Kindergarten Readiness: from 17% to 51%! One might view this result as extraordinary, and immediately want to scale the program. To understand why that choice is not prudent, consider what exactly we have learned from this research. If it is a typical social science experiment, then it has likely been conducted as an efficacy test: the 'best-case' test of the program is arm B versus the control, arm A. To understand why more information is necessary, we must consider the incentives that the researchers faced. Those incentives are set up to create a petri dish that provides results that gives the intervention its 'best shot,' or likewise the largest treatment effects. In this manner, we are answering the wrong question if we are attempting to provide policy advice. We are asking: can this idea work in the petri dish under the best-case situation rather than will this idea work at scale? This is the wrong question. We must not only do the efficacy test but also relevant tests of scale within the original discovery process. The economics of many situations demand such an approach."

William Nordhaus reflects on his career in economics in "Looking Backward, Looking Forward" (*Annual Review of Resource Economics*, 2024, 16: 1–20, https://www.annualreviews.org/content/journals/10.1146/annurev-resource-112223-091502). On the progress that has been made in analyzing climate change issues: "Because we know so much today, it is hard to remember how little we knew in 1970. Specialists thought the globe was cooling and that rising levels of particulates would further exacerbate cooling. The first study of the economics of climate change examined the impact of cooling, not warming. The only alternative to fossil fuels was thought to be nuclear power, which was viewed by many scientists with suspicion. All climate worries were based on modeling at that time, and only in the last two decades has the science been validated by observations. . . . Over the next two decades, I moved from model to model like a pilgrim trying to find the holy grail. The main goal was to fix the two major flaws in the IIASA [International Institute for Applied Systems Analysis] model: to develop a general-equilibrium framework and to develop the modules of the climate externality, especially developing monetized damage estimates. Accomplishing these two goals took nearly two decades, and what finally emerged was the DICE model (Dynamic Integrated model of Climate and the Economy). The first major study was rejected by economic journals but accepted and published in *Science* (Nordhaus 1992). I have always loved the name of the DICE model. It is easy to remember and visualize. DICE also conveys a shiver of risk and danger. It alludes to the Faustian bargain that we make as we continue down the path of unchecked climate change, on the Walpurgis Night of reveling in a world enflamed by fossil-fuel passions and ignoring how the devil of damages will drag us into a hellish future."

Symposia

For readers who would like more on merger issues than the symposium in this issue, the August 2024 issue of the *Review of Industrial Organization* contains a twelve-paper symposium on "The U.S. 2023 Merger Guidelines" (many of the papers are open access at https://link.springer.com/journal/11151/volumes-and-issues/65-1). Herbert Hovenkamp contributes "The 2023 Merger Guidelines: Law, Fact, and Method." From the abstract: "These Guidelines break some new ground that older Guidelines did not address, and make many positive contributions, which this paper spells out. They are also excessively nostalgic for a past era, however, and this may explain their propensity to treat empirical questions as issues of law: This is one way to insulate these Guidelines from further revision. The excessive reliance on one decision, *Brown Shoe*, is unfortunate—particularly since that decision has been so often repudiated, even by the Supreme Court itself. This paper pays particular attention to: the Guidelines' treatment of structural triggers and direct measures of competitive effects; their aggressive position on potential competition mergers; their willingness to weigh a 'trend' toward concentration as a factor; and their treatment of serial acquisitions. The Guidelines include a welcome new section on mergers involving multi-sided networks, although their view of networks is too one-sided; and the Guidelines also contain an expanded section on mergers with harmful effects on suppliers—including labor. The Guidelines' treatment of market definition is likely to lead to underenforcement because they define markets too broadly. Finally, the Guidelines could have made better use of recent retrospective studies—many of which would have provided further support for the substantive positions that the Guidelines take."

The Summer 2024 *Journal of Policy Analysis and Management* includes a Point/Counterpoint feature on reform of US health insurance (https://onlinelibrary.wiley.com/toc/15206688/2024/43/3). Liran Einav and Amy Finkelstein begin by offering "A blueprint for U.S. health insurance policy." "It contains two main elements. The first is universal coverage that is automatic, free to the patient, and basic. The second is the option—for those who want and can afford it—to purchase supplemental coverage in a well-functioning marketplace. We argued that we could thus fulfill our social contract without tackling the other multi-trillion-dollar elephant in the room: the problem of high and often inefficient healthcare spending. . . . In addition to the question of *what* will be covered, there is also a question of *how* it will be covered. The social contract is about providing essential medical care, not providing a high-end experience. There are many non-medical aspects of care that may be desirable without being essential. The ability to see the doctor of your choice at your preferred timing and location, for example, or semi-private hospital rooms. This would be substantially limited under basic coverage. Basic coverage would likewise involve longer wait times for non-urgent care than what people with private health insurance or Medicare are currently accustomed to. . . . Non-medical amenities would also be limited in the basic coverage, as they are in many other countries where hospitals offer a range of hotel-like amenities. In Singapore, for example, the basic coverage provides hospital care with eight beds in a room, a shared bathroom, and no air-conditioning. Patients can pay out of pocket to upgrade partway or all the way to the VIP treatment, which gets the patient a single room, with a private attached bath, television and, of course, air-conditioning, which is no small matter in Singapore's notoriously hot and humid climate. Australia's system is similar. . . . We estimate that about two thirds of Americans—those who are covered by Medicare or by private health insurance through an employer—would want to supplement beyond the basic." Jason Furman makes a case for a more incremental approach in "Starting health reform from here." "I worry that the basic benefits they propose would either grow to be much more than basic, coming at high expense, or would be subject to serious backlash. Rather than rely on just that one instrument to control costs, I would also include broader cost sharing, something we know works." The authors then respond to each other.

Addresses

In the annual Martin Feldstein Lecture at the National Bureau of Economic Research, Cecilia Elena Rouse suggests "Lessons for Economists from the Pandemic" (*NBER Reporter*, 2024, No. 3, https://www.nber.org/reporter/2024number3/lessons-economists-pandemic). Here's one of her lessons: "*Better calibrated economic policymaking will require much deeper investment in data and infrastructure.*" Rouse argues: "This lesson is based on the fact that federal data, computer, and human resource infrastructures were—and still are—not up to the task of delivering surgical and speedy support for the economy. . . . For example, the Paycheck Protection Program (PPP) provided uncollateralized and forgivable loans to small businesses (generally, those with fewer than 500 employees). These loans could officially be used only to retain workers (with several safe harbor provisions), meet payroll and health insurance costs, or make mortgage, lease, and utility payments. If these conditions were met and firms met their employment targets, the loans would be entirely forgiven after the pandemic. The Economic Injury Disaster Loan (EIDL) program provided low-interest-rate loans of up to $2 million, payable over up to 30 years. Loans also included the option to defer all payments during the first two years while businesses and nonprofits got back on their feet after the pandemic. And finally, the coverage and generosity of UI were expanded dramatically. Benefits were increased by $600 per week, and those not typically covered, such as gig workers and contractors, were made temporarily eligible. While it may have been 'good enough,' it was sloppy. . . . Waste and poor targeting were a problem. David Autor and his coauthors estimate that PPP loans cost between $169,000 and $258,000 per job-year saved, which is more than twice the average salary of these workers. They also estimate that more than two-thirds of the total outlays on the program accrued to business owners and shareholders rather than employees. Outright fraud was also a major issue. The Government Accountability Office (GAO) estimates that PPP fraud totaled about $64 billion out of a total of nearly $800 billion in loans . . . . Under EIDL, some borrowers claimed loans using falsified names or business details and often simply ran off with the cash. In the end, the GAO and the Small Business Administration estimate that EIDL fraud was even more pervasive than PPP fraud, in dollar terms—more than $136 billion.

UI fraud also skyrocketed during the pandemic; the GAO estimates that fraud may have cost anywhere from $55 to $135 billion."

In his da Vinci Medal Address, Donald MacKenzie considers "Material Political Economy" (*Technology and Culture*, July 2024, https://muse.jhu.edu/pub/1/article/933102). He focuses in part on the material side of high-frequency trading. "Just how fast is 'ultrafast'? Each year, the European futures exchange Eurex publishes data from which we can infer the response times of the fastest HFT algorithms. Eurex's 2023 measurements suggest a state-of-the-art response time (to a packet of market data that triggers a trading system's action) of 8 nanoseconds, or billionths of a second. In a nanosecond, the fastest physically possible signal, light in a vacuum, travels only around 30 cm, or roughly a foot. That is not simply a helpful yardstick of HFT's speed: getting messages to travel as close as possible to the speed of light in a vacuum is an important practical concern in HFT. Fiber-optic cable, for example, is not fast enough, because the refractive index of the glass at the core of such a cable slows laser-light signals to around two-thirds of light's speed in a vacuum. Where possible, therefore, HFT firms send trading data and orders by microwave, millimeter-wave, or laser-light signals transmitted through the atmosphere, where they travel almost as fast as in a vacuum . . . Since around 2010, furthermore, a conventional computer system, even if programmed in C++, is in many markets not fast enough for HFT. Trading algorithms are directly programmed into the hardware of the silicon chips known as field-programmable gate arrays (FPGAs) . . . There have been repeated rumors of firms moving beyond FPGAs to fully bespoke integrated circuits . . ."

Interviews

Orley Ashenfelter interviews Samuel Bowles "on his deep interest in the causes of inequality & his work to transform economics" ("The Work Goes On" podcast, posted October 7, 2024, https://soundcloud.com/theworkgoesonpodcast/a-conversation-with-samuel-bowles). For example, I did not know that Bowles attended school in a tent in India when he was eleven years old (his father was Ambassador to India at the time), nor that he worked for the government of Nigeria as a teacher in a remote area after graduating from college, nor that he offered economic advice to Martin Luther King, Jr. Bowles says: "I had the good fortune of being asked by Dr. Martin Luther King if I wouldn't give him some advice about economics and of course, I was thrilled. I knew Dr. King through anti-Vietnam War activity that he and I had engaged in at the time. And so, he said he would send me some questions and I said, 'well, I'll definitely get back to you.' I opened the envelope and here's a set of questions. And they're all about economics . . . I looked at these questions and I said, 'these are damn good questions. I don't have a clue how to answer these.' I didn't know where to look. They were empirical questions, but also conceptual ones. Imagine, a new successful PhD, and this could have been the high point in my life. This is why I studied economics so I could actually get into the fray and help out and make the world a better place. But that was kind of a shocker for me, and I decided that there was something wrong that I was actually teaching the grad students in micro. And I decided then and there that I was either going to leave economics, I considered that very seriously, or I would try to change it. And that's what I've been trying to do since."

Cardiff Garcia of the Economic Innovation Group interviews Paul Krugman in "How to Slay Economic Zombies" (The New Bazaar website, October 9, 2024, https://eig.org/newbazaar/how-to-slay-economic-zombies/). "[T]he modern forms of agglomeration are different from the ones that prevailed in the 19th century. I like to say that the models that I was writing down 30 years ago had this kind of steampunk feel. They all kind of were very much focused on manufacturing and on industrial clusters, and we all lavished attention on these great stories—like the detachable collar and cuff industry of Troy, New York, and that sort of thing. And which mostly have gone away in the United States, although not totally—they do exist to some extent even in manufacturing, but these days if you really want to find old style industrial clusters, you go to China. So if you actually look at—there's a variety of measures—but basically, in the United States, there was a lot of regional convergence, convergence in incomes, convergence in basically regions becoming more similar, from the 1920s up until about 1980. Then in 1980, they started pulling apart again, and you started to see metropolitan areas with highly educated workforces pulling in even more educated people, pulling in even more of the information economy—and stranding regions that didn't have those preconditions. . . . So we're back in a world where we have these extremely localized clusters. . . . Particularly for high skill, high pay workers, you need amenities. High tech workers are not going to move to someplace in the middle of the country, even if they have excellent internet access, because where are the good restaurants? Where are the live concerts? So in some ways the fact that we're rich enough that people can make decisions on that basis matter. So I think we're in some ways back to the kind of unequalizing development that we had in the late 19th century."

Joe Walker interviews Larry Summers on the subject of "artificial general intelligence" in "AGI and the Next Industrial Revolution" (The Joe Walker Podcast, October 22, 2024, https://josephnoelwalker.com/larry-summers-159/). "[T]he more I study history, the more I am struck that the major inflection points in history have to do with technology. I did a calculation not long ago, and I calculated that while only 7% of the people who've ever lived are alive right now, two-thirds of the GDP that's ever been produced by human beings was produced during my lifetime. And on reasonable projections, there could be three times as much produced in the next 50 years as there has been through all of human history to this point. . . . Of course, I think that this [AI] technology potentially has implications greater than any past technology, because fire doesn't make more fire, electricity doesn't make more electricity. But AI has the capacity to be self-improving."

Discussion Starters

The Stockholm Institute of Transition Economics (SITE) at the Stockholm School of Economics has tried to see through the smoke in its report "The Russian Economy in the Fog of War" (September 2024, https://www.hhs.se/en/about-us/news/site-publications/2024/russias-economic-imbalances/). "In terms of economic size, however, Russia is not a 'great power' with a GDP of around 2000 billion US dollars. That is about 1/10th of the combined GDP of the EU-27 (about 20 000 billion US dollars), or approximately the same size as the Nordic countries combined. The size of the US economy is about 27 000 billion US dollars or more than 13 times the Russian economy. Compared to other BRIC countries, Russia is behind Brazil (2200 billion US dollars), distanced with some margin by India (3600 billion US dollars), and only around 10 percent of the Chinese economy (17 800 billion US dollars). . . . [D]epending on what measure of GDP is used and the time-period that is included in the analysis, between 60 and 95 percent of Russia's GDP growth can be explained by changes in one exogeneous variable alone: the change in international oil prices. . . . The current state of the Russian economy according to official statistics is then covered by going through standard economic indicators such as GDP growth, inflation, monetary policy, fiscal policy, reserves, trade, the exchange rate and the financial sector. It then offers a critical analysis of this official view and questions the credibility of two key economic indicators, growth and inflation, where it is shown that inflation may be significantly higher than official numbers suggest and growth significantly lower."

Julian Morris and Ben Sperry offer "The Cost of Payments: A Review," (International Center for Law & Economics Working Paper, August 28, 2024, https://laweconcenter.org/resources/the-cost-of-payments-a-review/). "Atlanta's Mercedes-Benz Stadium in 2018 became the first major sports venue in the United States to switch to a fully cashless payment system. At the end of the new payment model's first year of operations, the stadium reported that wait times had fallen by 20 to 30 seconds and per-capita food and beverage sales had risen by 16%, while saving more than $350,000 in operating expenses. . . . In short, the evidence shows that, when all costs and all parties to a transaction are considered, electronic payments (debit cards, credit cards, and mobile payments) are more cost-effective than cash for most transactions. The main reason for this is that electronic payments enable consumers to spend more than they have in their wallet, which results in 'ticket lift' for merchants. Card rewards, including cashback and merchant-specific loyalty programs, further increase this ticket lift. In addition, 'tap-and-pay' contactless payments can reduce the time it takes to tender payment relative to cash, especially when cash payments are eliminated altogether. This increases throughput, improving the customer experience and reducing labor costs. Finally, electronic payments enable merchants to sell online, including for in-store pickup."

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to JEP articles, go to the AEA website: http://www.aeaweb.org. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of

Economic Perspectives

Winter 2025, Volume 39, Number 1

Symposia

Articles

Features

AMERICAN
ECONOMIC
ASSOCIATION