

The Journal of

Economic Perspectives

*A journal of the
American Economic Association*

Spring 2022

The Journal of **Economic Perspectives**

A journal of the American Economic Association

Editor

Heidi Williams, Stanford University

Coeditors

Erik Hurst, University of Chicago

Nina Pavcnik, Dartmouth College

Associate Editors

Gabriel Chodorow-Reich, Harvard University

David Deming, Harvard University

Andrea Eisfeldt, University of California at Los Angeles

Shawn Kantor, Florida State University

Eliana La Ferrara, Bocconi University

Camille Landais, London School of Economics

Amanda Pallais, Harvard University

Nancy Rose, Massachusetts Institute of Technology

Juan Carlos Serrato, Duke University

Charlie Sprenger, University of California, San Diego

Francesco Trebbi, University of California, Berkeley

Lise Vesterlund, University of Pittsburgh

Gianluca Violante, Princeton University

Ebonya Washington, Yale University

Managing Editor

Timothy Taylor

Assistant Managing Editor

Grace Aquilina

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 (®).

Copyright © 2022 by the American Economic Association; All Rights Reserved.

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.

Printed by LSC Communications, Owensville, Missouri, 65066, USA.

No responsibility for the views expressed by the authors in this journal is assumed by the editors or by the American Economic Association.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Spring 2022, Vol. 36, No. 2. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are \$24.00, \$34.00, or \$44.00 depending on income; for an additional \$15.00, you can receive this journal in print. The journal is freely available online. 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 U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 36 • Number 2 • Spring 2022

Symposia

Macro Policy in the Pandemic

- Christina D. Romer and David H. Romer, “A Social Insurance Perspective on Pandemic Fiscal Policy: Implications for Unemployment Insurance and Hazard Pay” 3
- Giulia Giupponi, Camille Landais, and Alice Lapeyre, “Should We Insure Workers or Jobs During Recessions?” 29
- David Autor, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz, “The \$800 Billion Paycheck Protection Program: Where Did the Money Go and Why Did It Go There?” 55

Economics of Slavery

- Trevon D. Logan, “American Enslavement and the Recovery of Black Economic History” 81
- William Darity Jr., A. Kirsten Mullen, and Marvin Slaughter, “The Cumulative Costs of Racism and the Bill for Black Reparations” 99
- Gavin Wright, “Slavery and the Rise of the Nineteenth-Century American Economy” 123

Childhood Interventions

- Anna Aizer, Hilary Hoynes, and Adriana Lleras-Muney, “Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children” 149
- Miriam Wüst, “Universal Early-Life Health Policies in the Nordic Countries” . . . 175
- Sarah Flood, Joel McMurry, Aaron Sojourner, and Matthew Wiswall, “Inequality in Early Care Experienced by US Children” 199
- Anthony Bald, Joseph J. Doyle Jr., Max Gross, and Brian A. Jacob, “Economics of Foster Care” 223

Features

- Carolina Alves, “Retrospectives: Joan Robinson on Karl Marx: ‘His Sense of Reality Is Far Stronger’” 247
- Timothy Taylor, “Recommendations for Further Reading” 265

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. 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

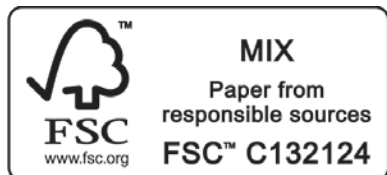
Advisory Board

Kerwin Charles, Yale University
Karen Dynan, Harvard University
Peter Henry, New York University
Marionette Holmes, Spelman College
Soumaya Keynes, *The Economist*
Trevon Logan, Ohio State University
Emily Oster, Brown University
Lucie Schmidt, Smith College
Dan Sichel, Wellesley College
Jonathan Skinner, Dartmouth College
Matt Taddy, Amazon.com, Inc.
David Wessel, Brookings Institution

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>. Annual dues for regular membership are \$24.00, \$34.00, or \$44.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

Copyright © 2022 by the American Economic Association. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation, including the name of the author. Copyrights for components of this work owned by others than AEA must be honored. Abstracting with credit is permitted. The author has the right to republish, post on servers, redistribute to lists, and use any component of this work in other works. For others to do so requires prior specific permission and/or a fee. Permissions may be requested from the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203; email: aeainfo@vanderbilt.edu.



EXECUTIVE COMMITTEE

Elected Officers and Members

President

CHRISTINA D. ROMER, University of California, Berkeley

President-elect

SUSAN C. ATHEY, Stanford University

Vice Presidents

DAVID H. AUTOR, Massachusetts Institute of Technology

CAROLINE M. HOXBY, Stanford University

Members

AMANDA BAYER, Swarthmore College

SANDRA E. BLACK, Columbia University

LISA D. COOK, Michigan State University

MELISSA S. KEARNEY, University of Maryland

EMI NAKAMURA, University of California, Berkeley

MELVIN STEPHENS, JR., University of Michigan

Ex Officio Members

DAVID CARD, University of California, Berkeley

Appointed Members

Editor, The American Economic Review

ESTHER DUFLO, Massachusetts Institute of Technology

Editor, The American Economic Review: Insights

AMY FINKELSTEIN, Massachusetts Institute of Technology

Editor, The Journal of Economic Literature

STEVEN N. DURLAUF, University of Chicago

Editor, The Journal of Economic Perspectives

HEIDI WILLIAMS, Stanford University

Editor, American Economic Journal: Applied Economics

BENJAMIN OLKEN, Massachusetts Institute of Technology

Editor, American Economic Journal: Economic Policy

ERZO F.P. LUTTMER, Dartmouth College

Editor, American Economic Journal: Macroeconomics

SIMON GILCHRIST, New York University

Editor, American Economic Journal: Microeconomics

LEEAT YARIV, Princeton University

Secretary-Treasurer

PETER L. ROUSSEAU, Vanderbilt University

OTHER OFFICERS

Director of AEA Publication Services

ELIZABETH R. BRAUNSTEIN

Counsel

LAUREN M. GAFFNEY, Bass, Berry & Sims PLC
Nashville, TN

ADMINISTRATORS

Director of Finance and Administration

BARBARA H. FISER

Convention Manager

GWYN LOFTIS

The Journal of
Economic Perspectives

Spring 2022, Volume 36, Number 2

Symposia

Macro Policy in the Pandemic

Christina D. Romer and David H. Romer, “A Social Insurance Perspective on Pandemic Fiscal Policy: Implications for Unemployment Insurance and Hazard Pay”

Giulia Giupponi, Camille Landais, and Alice Lapeyre, “Should We Insure Workers or Jobs During Recessions?”

David Autor, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz, “The \$800 Billion Paycheck Protection Program: Where Did the Money Go and Why Did It Go There?”

Economics of Slavery

Trevon D. Logan, “American Enslavement and the Recovery of Black Economic History”

William Darity Jr., A. Kirsten Mullen, and Marvin Slaughter, “The Cumulative Costs of Racism and the Bill for Black Reparations”

Gavin Wright, “Slavery and the Rise of the Nineteenth-Century American Economy”

Childhood Interventions

Anna Aizer, Hilary Hoynes, and Adriana Lleras-Muney, “Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children”

Miriam Wüst, “Universal Early-Life Health Policies in the Nordic Countries”

Sarah Flood, Joel McMurry, Aaron Sojourner, and Matthew Wiswall, “Inequality in Early Care Experienced by US Children”

Anthony Bald, Joseph J. Doyle Jr., Max Gross, and Brian A. Jacob, “Economics of Foster Care”

Features

Carolina Alves, “Retrospectives: Joan Robinson on Karl Marx: ‘His Sense of Reality Is Far Stronger’”

Timothy Taylor, “Recommendations for Further Reading”



A Social Insurance Perspective on Pandemic Fiscal Policy: Implications for Unemployment Insurance and Hazard Pay

Christina D. Romer and David H. Romer

C OVID-19 is a worldwide public health crisis unlike anything seen in the post-World War II era. The fiscal response has been similarly extraordinary. The United States has spent \$5.2 trillion on a wide range of recovery and support initiatives, and other countries have also spent unprecedented sums. The pros and cons of various measures have been discussed extensively. But, what has been largely missing from the fiscal policy discussion is an overarching analytical framework that takes into account the unique nature of a pandemic recession. This paper seeks to fill this gap.

We argue that a social insurance perspective is the appropriate way to understand and evaluate fiscal policy in a pandemic. Social insurance analysis shows how the government can use taxes and transfers to provide people with insurance that they would like to have, but that doesn't exist or that no one had contemplated needing. During a pandemic, workers in certain sectors face prolonged unemployment because their industries can't operate safely, while workers in other sectors remain relatively unscathed. Had workers foreseen this possibility, they would have liked to purchase insurance against the risk that their sector would be closed. The social insurance framework can show which types of government fiscal actions best approximate what a well-functioning insurance market would provide.

A social insurance perspective is more appropriate for designing and evaluating pandemic fiscal policy than simple aggregate-demand-based models. Conventional

■ *Christina D. Romer is Class of 1957-Garff B. Wilson Professor of Economics and David H. Romer is Herman Royer Professor in Political Economy, both at the University of California, Berkeley, Berkeley, California. Their email addresses are cromer@berkeley.edu and dromer@berkeley.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.3>.

Keynesian models of fiscal policy suggest that the way to deal with a recession is to increase aggregate demand quickly, and by enough to return output to its normal or potential level. And in this framework, it is not necessary for fiscal policy to closely target the workers or industries most affected by the recession. Raising aggregate demand anywhere will raise incomes and spending throughout the economy, and so help will eventually flow to those most affected.

These models and policy prescriptions don't hold in a pandemic recession. Because the virus thrives on human interaction (and hence on some types of economic activity), fiscal policy should not be aimed at quickly raising aggregate demand and attempting to return the economy to full employment. Doing so would make the pandemic worse and increase illness and deaths. Similarly, in a pandemic, some types of economic activity—such as in-restaurant dining and cruise travel—simply can't take place safely. As a result, broad stimulus measures like one-time payments or tax cuts can do little to put workers in those industries back to work. In essence, pandemic-related shutdowns of certain sectors short-circuit the usual Keynesian multiplier effect.

We begin our analysis by developing the social insurance framework. We describe a simple model with one sector that is affected by the pandemic and one that isn't. We show that optimal policy involves the government taxing those in the unaffected sector and providing income support for those in the sector that is shut. We then discuss several enhancements of the framework that yield richer understanding and more nuanced policy prescriptions. For example, adding a third sector where essential workers remain employed, but face greater health risks because of the nature of their jobs, suggests that government-provided hazard pay is appropriate. Or, incorporating notions of fairness or difficulty in identifying which workers remain able to work implies that optimal policy includes a role for general stimulus.

Armed with the framework, we examine two types of pandemic fiscal policy in detail: one that has been used extensively—unemployment insurance—and one that's hardly been used at all—hazard pay. We discuss in more depth the implications of a social insurance perspective for the usefulness of such policies and how they should be structured. We also discuss some of the practical issues around designing these fiscal actions for use in a pandemic, and examine how the implications of a social insurance perspective compare with what was actually enacted or proposed during the pandemic.

Finally, in the conclusion we consider the broader applicability of the social insurance perspective on fiscal policy. We argue that the insights likely carry over to a wide range of situations other than the current pandemic.

Some previous authors have also suggested a social insurance perspective on the appropriate policy response to the pandemic, and the baseline case of the next section draws heavily on that work.¹ Our contributions are in the ways we go beyond

¹In a contribution early in the pandemic, Milne (2020) argues that the idea of “retrospective insurance” provides a valuable way of thinking about important aspects of the appropriate policy response. Similarly,

the baseline case and in evaluating actual and proposed policies from a social insurance perspective.

A Social Insurance Perspective

The basic logic of a social insurance perspective on the fiscal policy response to a pandemic is straightforward. We therefore present a verbal description here, and leave the more formal presentation to the online Appendix available with this article at the *JEP* website.

A Baseline Case

To see many of the key implications of the social insurance perspective, we start with a very stylized case that builds on Guerrieri et al. (2020) and Woodford (2020). We then consider the implications of relaxing some of its assumptions to address additional issues.

Think of a competitive, one-period economy with many identical individuals. There are two sectors, *A* and *B*. Other than producing different outputs, the two sectors are identical. Each individual obtains utility from their consumption of both sectors' outputs. Consumption of one sector's output doesn't affect utility from consuming the other sector's (that is, utility is "additively separable" over consumption of the two outputs). Although this assumption of course isn't exactly correct, the idea that, for example, whether one is able to go the dentist or take an airplane flight doesn't affect the marginal utility of groceries or clothes is arguably a reasonable approximation. We also assume that individuals get disutility if they work. For simplicity, work is a 0–1 variable—an individual is either working or they aren't.

We think of a pandemic as an event that makes production in one sector impossible; this could be either because producing the sector's output is unsafe or because consuming it is. To preserve the symmetry between the sectors, we assume they have the same probability of being shut down by a pandemic.

Finally, individuals must decide which sector they want to work in before they learn whether a pandemic will occur. Allowing for some mobility after the pandemic arises mitigates the fall in output in a pandemic, but otherwise has little impact on the messages of the baseline case.

in an informal early contribution, Saez and Zucman (2020) propose full government replacement of lost income for workers and businesses as a form of social insurance. Our framework builds most closely on Woodford (2020) and, especially, Guerrieri et al. (2020). Both papers consider multi-sector models where one sector is forced to shut down because of a pandemic, and both consider social insurance policies. And like us, Woodford uses the hypothetical case where there are perfectly functioning markets for "pandemic insurance" as a benchmark. These authors' main interests, however, are different from ours: the focus of Guerrieri et al. is on the conditions under which a pandemic—which is fundamentally a shock to aggregate supply—can lead to an aggregate demand shortfall (an issue we discuss below), while Woodford's focus is on the consequences of the structure of linkages among sectors.

The equilibrium of this stylized economy isn't hard to describe. Consider first what happens when individuals cannot insure against a pandemic. This could occur because people simply hadn't contemplated the possibility of a pandemic, or because in practice the difficulty of spelling out exactly what constitutes a pandemic makes the cost of writing insurance contracts prohibitive. The symmetry of the model implies that half of individuals are in each sector. In the absence of a pandemic, everyone earns the same amount, each sector produces the same amount, and each individual consumes the same amount of each sector's output. If a pandemic shuts one sector, the individuals who were working there earn no income, and so have no consumption (recall that the economy lasts for only one period and workers can't switch sectors in a pandemic). The individuals in the sector that stays open continue to work, but they now spend all their income on that sector's output rather than splitting it between the two sectors.

In normal times, because everyone's consumption is the same, everyone's marginal utility of consumption is the same. But in a pandemic, the marginal utility of consumption (of the output of the sector that remains open) of those who remain employed is lower than normal, while the marginal utility of those who become unemployed is higher than normal (probably greatly so, because their consumption falls to zero). Thus, the marginal utility of sector-*A* workers is higher than that of sector-*B* workers if there's a sector-*A* pandemic, but lower if there's a sector-*B* pandemic. This variation in relative marginal utilities implies that from an *ex ante* perspective—that is, before it's known whether there will be a pandemic—the outcome is Pareto inefficient. Measures that would shift resources from low-marginal-utility individuals to high-marginal-utility individuals in a pandemic would raise everyone's *ex ante* expected utility.

The efficient allocation can be achieved without any government action if there are not only competitive markets for output and labor in each sector after it is known whether there will be a pandemic, but also competitive markets for “pandemic insurance” before it's known whether there will be a pandemic. With these markets in place, the outcome in the absence of a pandemic is the same as before. But now individuals in each sector purchase insurance against the possibility of a pandemic hitting their sector (with the market clearing because the individuals in the other sector are willing to sell such insurance). In the event of a pandemic, every individual consumes the same amount of the output of the sector that stays open as they do in the absence of a pandemic. Marginal utility is always equal across all individuals, and so the allocation is efficient.²

The final step is to return to the case where there is no insurance and consider how the government can implement the efficient allocation—that is, the one that

²Another market-based way of achieving the efficient allocation is through trade in competitive markets in the full set of “state-contingent” (or “Arrow-Debreu”) commodities (output and labor in each sector in each state of the world—no pandemic, a sector-*A* pandemic, and a sector-*B* pandemic) before it's known whether there will be a pandemic. This produces the same outcomes as with competitive markets for pandemic insurance. Although this set-up is obviously more abstract and even less realistic than the possibility of pandemic insurance, it can be useful for clarifying outcomes and their welfare properties.

would occur with perfectly functioning insurance markets. In this case, the public sector is providing the insurance individuals would obtain for themselves if markets functioned perfectly, so it's natural to describe the policy as social insurance.

The policy that reproduces the efficient outcome is simple. In the absence of a pandemic, the government takes no action. But if there is a pandemic, it taxes the individuals in the sector that is still open half their income and transfers the proceeds to the individuals in the sector that is shut. The result is that each individual's after-tax-and-transfer income is half their usual income, and everyone's consumption of the output of the sector that stays open is the same as in normal times.

Thus, the policy that replicates what would happen with perfect insurance markets is one of targeted transfers: the government makes transfers to individuals who can't work because of the pandemic. Moreover, there is no "stimulus" in the optimal policy: what the government pays out in transfers equals what it takes in from taxes, and output and employment are the same as they would be without insurance markets or government intervention (as also observed by Woodford 2020). In addition, although individuals who can't work because of the pandemic obtain the same after-tax-and-transfer income (and the same consumption) as those who stay employed, they don't attain their usual income. Instead, they only get what they normally spend on the output of the non-pandemic sector. In that sense, insurance is less than complete.

Finally, a simple way of thinking about the government's policy is that it taxes the individuals in the non-pandemic sector what they would normally spend on the pandemic sector's output and gives the proceeds to the individuals in the pandemic sector. This allows them to maintain their normal spending on the output of the non-pandemic sector.

Although this baseline case shows some key messages of the social insurance perspective, it omits some important issues. We therefore turn to extensions. Table 1 summarizes their key features and implications.

Incentives and Fairness

In the efficient allocation of the baseline case, the individuals who were working in the sector that shuts down are actually better off than those in the sector that stays open: everyone has the same consumption, but only the individuals in the sector that's open have the disutility of working. There are two reasons that implementing that allocation through government social insurance might not be workable. The first involves incentives. In the model, it's evident who should continue working in the pandemic—everyone working in the sector that isn't shut by the pandemic. In practice, however, the situation is more complicated. Individuals differ in their nonwork situations (such as their childcare needs and preexisting health conditions), in their attitudes toward health risks, and in how easily they can change sectors in a pandemic. Similarly, it often isn't obvious which firms (or which parts of a firm) should continue operating in a pandemic. As a result, for many workers it is not possible to tell whether the optimal outcome involves their continuing to

Table 1

Variations on the Baseline Case

	<i>Variant</i>	<i>Key implication</i>
1.	Incentive or fairness considerations	Unemployed workers have lower consumption than the employed
2.	A high-risk essential sector (plus incentive or fairness considerations)	Hazard pay for high-risk essential workers
3.	Heterogeneous incomes, possibilities for self-insurance, and costs of social insurance	Higher replacement rates of lost income for lower-income workers
4.	Multiple periods, barriers to borrowing, and a force causing the unemployed to have lower consumption than the employed	Aggregate demand falls by more than aggregate supply, implying a role for aggregate demand stimulus

work. Second, public perceptions of fairness don't align perfectly with allocations that would prevail under complete markets (see, for example, Weinzierl 2014), and so an allocation that makes the unemployed better off than the employed may not be politically feasible or may be viewed as undesirable.

We capture these considerations by assuming the government has no information about who should optimally be working in a pandemic and who shouldn't. As a result, the individuals who continue working must be at least as well off as those who don't; otherwise, no one would work. Since working involves disutility, it follows that the employed must have greater consumption than the unemployed. Total output (and hence total consumption) is the same as without this constraint. Thus the employed consume more than they do without the constraint, and the unemployed consume less. The "replacement rate" from social insurance is therefore lower than in the baseline case. The other key implications of the baseline case continue to hold: the optimal policy is targeted transfers, and there's no need for stimulus.³

High-Risk Essential Workers and Hazard Pay

During a pandemic, there are some jobs—notably in healthcare—where it's socially desirable for workers to continue working even though it involves an elevated risk of illness. To capture this, suppose there are three sectors rather than two. When there's a pandemic, as before one sector shuts down and one is unaffected. But now

³The online Appendix also discusses the cases where the government has some but less than full information about who should be working in a pandemic, and where some individuals can switch sectors in a pandemic but the government has limited information about who can switch easily. For the most part, these extensions don't change our main messages. However, we find that labor mobility with limited government information not only leads to a smaller fall in output in a pandemic than in the baseline case, but also provides a reason the government might want to provide a "moving bonus" to workers who switch sectors in a pandemic (somewhat analogous to the possibility of hazard pay discussed below).

a third sector is essential and high-risk: the workers are particularly valuable during a pandemic but must interact with others, and so face greater chances of illness. As a result, both the utility of consuming the sector's output and the disutility of working in this sector are higher than normal. We continue to assume that workers must decide which sector to be in before knowing whether there will be a pandemic. Finally, to maintain the symmetry across sectors (which makes the implications particularly clear), we assume that in a pandemic each sector is equally likely to be the one that shuts and each is equally likely to be the essential high-risk one.

In the absence of incentive or fairness considerations, the efficient policy with a high-risk sector is very similar to that of the baseline case. Workers are equally divided among the sectors. In a pandemic, each individual's consumption of the outputs of the two sectors that stay open is unchanged (while their consumption of the output of the sector that shuts of course falls to zero). As before, the efficient allocation can be obtained by a social insurance program of targeted transfers to the unemployed.

With this policy, however, workers in the high-risk sector are the worst off, individuals who had been working in the sector that shuts are the best off, and workers in the sector that stays open but is not high-risk are in between. As a result, the presence of an essential high-risk sector has more interesting implications if allocations that make the employed worse off than the unemployed aren't possible. With this constraint, as before, workers in the sector that stays open but is not hazardous must have higher consumption than the unemployed. More importantly, because working in the high-risk sector involves greater disutility than working in the other open sector, the consumption of workers in the high-risk sector needs to be even greater than that of workers in the other open sector.

This allocation can't be implemented just by targeted transfers to the unemployed; there must also be transfers to workers in the high-risk sector. This implication corresponds to the idea of "hazard pay": workers in the sector that is essential but hazardous are compensated for the additional risks they face by continuing to work.

Heterogeneous Incomes and Self-Insurance

In the baseline case, individuals in the sector that shuts down are fully compensated for the loss of the income they would use to purchase the output of the other sector. It follows that if earnings are heterogeneous, individuals with much higher non-pandemic earnings who lose their jobs receive much larger social insurance payments than lower-earning individuals. And although the situation with incentive or fairness considerations is somewhat more complicated, the result that optimal social insurance payments rise strongly with non-pandemic earnings carries over to that case.

This implication of our framework doesn't correspond well with intuition or with what is done in almost all social insurance programs. For example, although unemployment insurance payments are an increasing function of prior earnings, they're normally capped at a relatively low level, and that policy continued during the pandemic. Similarly, the various rounds of direct payments to individuals

excluded those with high incomes, and there were limits on the amount of an individual's earnings that could be paid using funds from the Paycheck Protection Program (where the government made forgivable loans to small businesses hurt by the pandemic).

This gap between our framework's implications and actual policy reflects, at least in part, the framework's omission of two important factors. One is that insurance (whether provided privately or through the government) involves administrative and related costs, which reduces the optimal amount of insurance. The other is that individuals make decisions over multiple periods. As a result, they have some ability to self-insure through their saving and borrowing, and this ability is almost certainly greater for individuals with higher incomes.

The availability of self-insurance reduces the amount of insurance the government should provide when that insurance is costly, and the reduction is greater when an individual's ability to self-insure is greater. Specifically, suppose (realistically) that higher-income individuals are more able to self-insure, in the sense that their consumption falls by a smaller proportion if they become unemployed in a pandemic and there's no social insurance. Then, as we show in the online Appendix, under reasonable assumptions pandemic social insurance payments rise less than proportionately with individuals' normal incomes. Indeed, if self-insurance is sufficiently strong at high enough incomes, optimal pandemic social insurance doesn't provide very high-income individuals with any payments at all.

The Possibility of an Aggregate Demand Shortfall and a Need for Stimulus

A pandemic is a disruption to the economy's ability to produce, so it's natural to describe it as a shock to aggregate supply. However, as argued informally by Rowe (2020) early in the pandemic and formalized soon after by Guerrieri et al. (2020), a pandemic that disproportionately affects some sectors can lead endogenously to a fall in aggregate demand larger than the fall in aggregate supply. This result holds in a natural multi-period version of our framework.

Following Guerrieri et al. (2020), we say aggregate demand falls by more than aggregate supply in a pandemic if with the previous real interest rate and the optimal social insurance policy, demand for output is less than the efficient level of output (which is, of course, less than normal output because of the need to shut down some of the economy). If this occurs, obtaining the efficient level of output requires some source of additional aggregate demand, which could come either from a reduction in the real interest rate or from fiscal stimulus.

Whether a pandemic reduces demand by more than supply isn't immediately clear. On the one hand, everyone's consumption falls in a pandemic. In a multi-period setting, this effect tends to make individuals want to borrow against their future income to smooth their consumption. This acts to raise demand relative to supply, and so works in the direction of raising the equilibrium real interest rate. On the other hand, the range of outputs that are available to consume falls in a pandemic because of the shutdowns. This makes consumption less attractive than usual, which works in the other direction.

It turns out that in the most natural multi-period extension of our baseline case, these two effects exactly balance. With optimal social insurance in the baseline case, each individual's consumption of the output of the sector that stays open is the same as in normal times. Together with the assumption that consumption of one sector's output doesn't affect utility from consuming the other sector's, this implies that each individual's marginal utility is the same as normal. As a result, in a multi-period setting, the same real interest rate that makes individuals want to buy the usual level of output in normal times makes them want to buy the efficient level of output in a pandemic. Thus, there is no force causing the equilibrium real interest rate to either rise or fall in a pandemic.

Crucially, however, this result fails if social insurance doesn't equalize consumption across individuals in a pandemic (as occurs with incentive or fairness considerations or with costly insurance). In this case, since total output of the sector that stays open is the same as normal, in the optimal allocation some individuals consume more of the sector's output than normal and some consume less. Those consuming more want to save, which reduces demand, while those consuming less want to borrow, which raises demand. If there are smaller barriers to saving than to borrowing, which is surely realistic, the saving effect dominates the borrowing effect, and so there's an aggregate demand shortfall. Thus, a pandemic is likely to lead to a need for aggregate demand stimulus.

Guerrieri et al. (2020) explore these issues in much more depth. For our purposes, the key message is that a need for stimulus in a pandemic isn't just possible but plausible despite the fall in the economy's safe capacity. One implication is that in thinking about possible social insurance policies in a pandemic, it's appropriate to be concerned about whether they raise aggregate demand in addition to providing insurance.⁴

Implications for Unemployment Insurance in the Pandemic

A key implication of the social insurance framework is that benefits should be targeted to those who suffer direct economic harms from the pandemic. Because being unable to work is a powerful indicator of being harmed economically, it follows that unemployment insurance should be a central component of pandemic fiscal policy. We therefore turn to what the social insurance perspective implies about how unemployment insurance should be structured in a pandemic, and compare those prescriptions with what was actually done in the United States.

⁴The online Appendix discusses two issues concerning pandemic fiscal policy that fit in the social insurance framework but that we don't consider in the text: "forgiveness" (that is, policies that eliminate either payment obligations, such as student loan payments, or outstanding debts, such as student loan balances) and aid to state and local governments. It also describes some extensions that have little effect on the main messages of our analysis.

Coverage

A first issue is who should be eligible for unemployment insurance. In practice, eligibility in normal times is restricted in various ways. Among the most important are that workers must have earned at least a certain amount over some length of time before becoming unemployed (with earnings from self-employment not counting toward this requirement); they must not have left their previous job voluntarily; and they must be either actively seeking work or awaiting recall to their previous job. Yet if individuals could purchase pandemic insurance, they would want to insure themselves against being unable to work in a pandemic regardless of whether they satisfied these criteria. Thus, the social insurance framework implies that unemployment insurance should be broadly available to individuals who aren't working because of a pandemic.

Actual policy in the United States largely followed this implication of our framework. In response to the pandemic, policymakers increased the length of time individuals could receive benefits and relaxed search requirements, waiting periods, and some eligibility rules for regular unemployment insurance. They also enacted the Pandemic Unemployment Assistance program, which extended coverage to many previously uncovered workers, including the self-employed, gig workers, and workers who quit their jobs because of childcare needs resulting from the pandemic (Cajner et al. 2020). Finally, as discussed in more detail below, they added a fixed weekly amount to benefits (including those through Pandemic Unemployment Assistance) for much of the pandemic.

As Cajner et al. (2020) stress, a host of factors—ranging from fraudulent claims to idiosyncratic biweekly reporting in some large states—make it very challenging to determine the number of individuals receiving benefits during the pandemic. Nonetheless, it's clear that coverage rose broadly in line with job losses. Employment (as measured by the household survey, and so including the self-employed) fell by 24.7 million from February to April 2020.⁵ Cajner et al.'s preferred measure of continued regular unemployment insurance claims rose by 19.7 million from its pre-pandemic level to its peak. Determining the number legitimately receiving Pandemic Unemployment Assistance benefits is even harder, but Cajner et al.'s results indicate it was at least several million.

At the same time, unemployment insurance coverage wasn't complete. Some of the pandemic unemployed, such as most new entrants to the labor force who could not find a job immediately, were not eligible for any unemployment insurance. In addition, the conjunction of the enormous crush of claims, limited processing resources, and antiquated administrative systems in many states led to long delays in processing claims (Bitler, Hoynes, and Schanzenbach 2020; Cajner et al. 2020). Nonetheless, there was unusually broad coverage of the unemployed in the pandemic.

⁵Because the impact of the pandemic swamped the effects of normal seasonal variation, the numbers in this paragraph are not seasonally adjusted.

Duration

A second issue that is not directly addressed by the one-period framework is how long pandemic-related unemployment insurance should last. Again, however, the logic of the social insurance perspective is clear. If individuals could purchase pandemic insurance, they would want to be insured for the duration of the pandemic. Thus pandemic-related unemployment insurance should be available for however long the pandemic lasts.

Actual US policy departed considerably from this implication of the social insurance framework. Policymakers did increase the number of weeks that individuals could receive benefits. However, throughout the pandemic, the various extensions of duration, expansions of coverage, and increases in benefits were tied to calendar time rather than to metrics of the course of the pandemic or the state of the economy. Moreover, there was often great uncertainty about whether the emergency measures would be extended or allowed to lapse, which is not what beneficiaries would desire from a social insurance program.

Consider, for example, the Pandemic Unemployment Assistance program. The original Coronavirus Aid, Relief, and Economic Security (CARES) Act, enacted on March 27, 2020, provided 39 weeks of benefits under the program and set its end date as December 31, 2020. The program was continued at the last minute through mid-March 2021 by the Consolidated Appropriations Act, 2021, enacted December 27, 2020, which also provided an additional eleven weeks of benefits. The program was extended again (through September 6, 2021) by the American Rescue Plan Act of 2021, enacted on March 11, 2021—just a few days before the program was to end.

Similarly, the additions to benefits fluctuated greatly in ways largely unrelated to the severity of the pandemic. The CARES Act provided an extra \$600 per week of unemployment insurance benefits through July 31, 2020, at which point the supplement was allowed to lapse. The Lost Wages Assistance Program through the Federal Emergency Management Agency then provided \$300 per week of additional benefits (or, in a few states, \$400 per week) for roughly six weeks for many workers, though workers receiving low benefits were excluded. Additional benefits then lapsed entirely through the end of 2020. The Consolidated Appropriations Act, 2021 provided \$300 per week in additional benefits from January 1, 2021, through March 14, 2021. Finally, the American Rescue Plan Act of 2021 extended the \$300 per week through September 6, 2021.

Replacement Rates

An issue for which it's much harder to determine the exact implications of the social insurance framework is how generous unemployment benefits in a pandemic should be and how that generosity should vary with prior income. In our baseline case, individuals who remain employed are taxed to finance transfers to the unemployed, and insurance payments to the unemployed are less than their prior incomes. In practice, however, additional unemployment insurance in the pandemic has been financed in the short term by government borrowing. In this case, as we describe in the online Appendix, the natural modification of our

baseline case implies that optimal social insurance fully replaces the incomes of individuals who can't work because of the pandemic.

The various extensions of our baseline case discussed in the previous section, however, suggest two reasons for less than 100 percent replacement. First, there may be incentive or fairness considerations that imply the unemployed shouldn't be better off than the employed. Because the unemployed don't have the disutility of working, in a one-period setting this requires that they have lower incomes than the employed. As various authors have pointed out, however, things are more complicated in a dynamic setting. Someone who is unemployed may face a long and challenging job search when the pandemic ends. Thus it's not clear that replacement rates need to be much (or even at all) below 100 percent during the pandemic to prevent individuals preferring unemployment to employment (for example, Ganong, Noel, and Vavra 2020).

Second, and more importantly, providing social insurance involves costs beyond the insurance payments. Those costs include administrative expenses, the distortions caused by the incentive effects of the insurance, and the distortions from raising the additional revenues at some point to satisfy the government's intertemporal budget constraint. These costs imply that even in the absence of incentive or fairness issues, optimal social insurance should leave the consumption of the unemployed somewhat below that of the employed. In the online Appendix, we argue that 10 to 15 percent is a plausible ballpark figure for the size of the shortfall.

How much social insurance is needed to achieve this level of consumption for unemployed workers depends crucially on the possibility and cost of self-insurance. In particular, to make more precise statements about optimal replacement rates and how they should vary with prior income, we would want two types of information. The first is a comparison of the consumption of unemployed and otherwise similar employed individuals in the absence of unemployment insurance, as a function of their income. This would be informative about the extent of self-insurance. The second is evidence about how the unemployed finance their consumption in the absence of unemployment insurance. This would be informative about the cost of self-insurance. To the extent individuals use types of self-insurance that are more costly than government-provided social insurance (as might be true of borrowing at very high interest rates or selling highly illiquid possessions), it's optimal for the government to provide the insurance. On the other hand, if sufficiently high-income individuals can do enough self-insurance at a cost less than that of the government to keep their consumption from falling very much if they become unemployed, optimal social insurance doesn't provide them with any social insurance payments.

Ganong and Noel (2019) examine the consumption behavior of unemployment insurance recipients before the pandemic. In a sample where recipients are probably financially healthier than typical beneficiaries, they find that after recipients exhaust their benefits, their consumption is on average 20 to 30 percent below what it was when they were employed. They also find that the falls average 30 to 40 percent among those in the highest third of the ratio of benefits to prior income and in the lowest third of assets relative to prior spending (both which tend to be

associated with lower incomes). Farrell et al. (2020) and Ganong et al. (2021) find that delays in benefits early in the pandemic also resulted in large reductions in consumption for the unemployed.

Another type of evidence about these issues comes from short-run marginal propensities to consume (MPCs) out of stimulus payments during the pandemic. A high MPC is suggestive of current marginal utility being high relative to future marginal utility, and thus of short-run financial distress. Karger and Rajan (2021) and Baker et al. (2021) find large MPCs from the stimulus payments. More importantly for our purposes, they find that it varied substantially with income. For example, Baker et al.'s point estimates (from their Table 5, Column 1) imply an MPC over three months of 0.66 at a monthly income (net of withholding) of \$500 and just 0.15 at a monthly income of \$5000. This evidence points to greater economic stress during the pandemic among low-income individuals.

Taken together, these studies indicate that unemployed workers may have limited ability to self-insure, and that this ability may be substantially smaller among lower-income workers. This suggests that the replacement rate for unemployment benefits may need to be fairly substantial, though clearly less than 100 percent, to result in a loss of consumption in the 10 to 15 percent range. It also suggests that replacement rates should decline as prior income rises. However, the existing evidence is not enough to pin down optimal replacement rates precisely.

Even though we are unable to say what exactly replacement rates from unemployment insurance during a pandemic should be, it is clear that actual replacement rates have differed sharply from the prescriptions of a social insurance perspective. Ganong, Noel, and Vavra (2020) show that the flat \$600 per week of additional benefits raised replacement rates to well over 100 percent for most workers.

There appear to have been two forces behind the policies involving greater than 100 percent replacement. One is the pursuit of other objectives, especially redistribution toward lower-wage workers and aggregate demand stimulus. The other is idiosyncratic factors: Ganong, Noel, and Vavra (2020) report that an overestimate of the average wage of workers who would lose their jobs led policymakers to underestimate the impact of the \$600 weekly adjustment on replacement rates, and that the very limited capacities of state unemployment insurance systems led policymakers to adopt the fixed supplement rather than more complicated additions to benefits.

The later iterations of benefit increases, which were in the range of \$300 per week, likely reduced replacement rate to below 100 percent for many workers, and so are more consistent with the social insurance perspective. Also, regular unemployment benefits continued to phase out at high incomes, which is consistent with the social insurance framework given the evidence that self-insurance is easier for high-income unemployed workers.

Unemployment Insurance Payments as Demand Stimulus

As we have discussed, even with optimal social insurance, a pandemic can lead to a shortfall of aggregate demand from what is needed to yield the desirable level of output. Thus, even though the central motivation for broad unemployment

insurance in a pandemic is social insurance, it is valuable to know whether unemployment insurance is also an effective source of aggregate demand stimulus.

Three types of evidence suggest that it is. First, the social insurance framework implies that unemployment insurance should be targeted to individuals whose marginal utility from consumption would otherwise be temporarily high. Such individuals would be expected to devote a large fraction of a marginal increase in resources to current consumption.

Second, the evidence from before the pandemic points to a high marginal propensity to consume out of unemployment insurance benefits, and thus to it being effective stimulus. Ganong and Noel (2019) report a very conservative lower bound for the MPC out of unemployment insurance benefits of 0.27, an upper bound of 0.83, and a point estimate under reasonable assumptions of 0.77. In a survey of work in this area, Chodorow-Reich and Coglianesi (2019) suggest an MPC of about 0.35 over one month and 0.55 over a year.

Third, the evidence from the pandemic also points to high marginal propensities to consume out of unemployment insurance benefits. Most notably, Ganong et al. (2021) estimate an MPC out of benefits (both regular and the \$600 supplement) between 0.29 and 0.43 over one month, and between 0.62 and 0.69 over six months.

One wouldn't expect the large marginal propensities to consume to apply regardless of the level of benefits, however. If benefits are sufficiently high that the marginal utility of current consumption for the unemployed is driven down to its normal level, a large fraction of any additional benefits is likely to be saved. The evidence doesn't clearly support this prediction, however. Ganong et al. (2021) find that because of the high replacement rates, the unemployed had both unusually high consumption and large increases in their stock of savings during the period when benefits included the \$600 supplement. Furthermore, despite the higher saving, they estimate an MPC out of the subsequent \$300 supplement that is very similar to the MPC out of the initial benefits. Thus, the issue of under what circumstances MPCs are high remains open.

Implications for Hazard Pay in the Pandemic

As discussed in the first section, a novel implication of the social insurance perspective is that in some cases, the government should provide hazard pay to high-risk essential workers during a pandemic. This implication results from extending the baseline model to include a third sector where workers doing very socially valuable jobs face greater risk of illness, and hence have greater disutility of work during a pandemic. The inclusion of this third sector, combined with either notions of fairness or difficulty in determining who should be working, leads to paying such at-risk workers a bonus. Intuitively, the need for hazard pay comes from seeking to make sure workers employed in the riskier sector aren't worse off than workers whose sector had to shut down. This section investigates this implication of the

social insurance perspective on fiscal policy in more detail. It also considers some of the many practical issues involved in the design and implementation of an actual program.

Government-funded hazard pay was considered during the pandemic and actually implemented on a very limited basis. These programs provide a useful baseline for discussion. The 2020 HEROES Act, which passed by the House of Representatives in May 2020, contained a \$200 billion hazard pay program (US Congress 2020). The proposed program, which did not make it into final legislation, would have paid an extra \$13 per hour for the period from January 27, 2020 (the first day of the declared public health emergency) until 60 days after the last day of the emergency. The definition of who was potentially eligible was very broad. Qualifying work was defined as work “not performed while teleworking from a residence” and involving “regular in-person interactions with . . . patients; . . . the public; . . . or coworkers” in a number of specific areas (US Congress 2020). Actual eligibility determination required an application by the employer to the Treasury Department. If the application was approved, the employer would then receive a grant and be responsible for providing the premium pay to workers.

An example of a hazard pay program that was actually implemented is the COVID-19 Pennsylvania Hazard Pay Grant (Pennsylvania Department of Community and Economic Development 2020), enacted in May 2020. The employees receiving the hazard pay needed to interact with others and couldn’t be teleworking from home. The list of eligible industries was decidedly shorter than the HEROES Act program. The amount of hazard pay was \$3 per hour, up to a total of \$1200. Like the HEROES Act, the Pennsylvania program required employers to apply for the grants. Employers could only apply for grants covering up to 500 full-time-equivalent employees per location. The budget for the program was \$50 million.

Who Should Receive Hazard Pay?

Whether some workers should receive hazard pay depends crucially on differences in risk across occupations during the pandemic. A number of studies have tried to investigate this issue. Their findings are summarized in Table 2.

One of the most detailed studies combines comprehensive health records and occupation data to form a complete individual-level national database for Norway (Magnusson et al. 2021). The study then identified occupations with direct contact with children, students, patients, and customers, and compared COVID-19 infection rates for these occupations with those of all other working-age adults. The study found that after controlling for age and sex, healthcare workers and transit workers had approximately three times the risk of contracting COVID-19 during the first wave of infection (February 26–July 17, 2020) (Magnusson et al. 2021, Figure 3). Interestingly, in the second wave (July 18–December 18, 2020), healthcare workers no longer had significantly elevated COVID-19 risk. Instead, the occupations with the highest odds ratios of infection were food service workers (food counter attendants, bartenders, and waiters), transit workers, and cleaners (p. 8).

Table 2

Summary of Studies of Occupational Risk of COVID-19

<i>Study</i>	<i>Sample</i>	<i>Riskiest occupations</i>	<i>Degree of elevated risk</i>
Outcome Variable: COVID-19 Infection			
1. Magnusson et al. (2021)	All working-age Norwegians	1st wave: healthcare, public transit workers 2nd wave: food service, public transit workers, cleaners	Approximately 3 times all other working-age adults Approximately 1.5–2 times all other working-age adults
2. Mutambudzi et al. (2021)	120,075 UK Biobank participants	Healthcare workers Social care workers	7.43-fold increase in risk 2.46-fold increase in risk
3. Song et al. (2021)	400,000 primary policy holders of a major insurer in PA	Essential workers as revealed by whether their industry was shut down	53% increase in risk relative to non-essential workers
Outcome Variable: COVID-19 Death			
4. Billingsley et al. (2020)	All deaths in Sweden 3/5 to 5/7/2020	Taxi and bus drivers Service sector	3.7 times the risk of IT techs. 2.2 times the risk of IT techs. (not statistically significant)
5. Chen et al. (2021)	Deaths in CA from 1/2016 to 11/2020 of people 18–65	Food or agriculture Transportation or logistics Manufacturing Facilities	1.39 times expected deaths 1.31 times expected deaths 1.24 times expected deaths 1.23 times expected deaths

A study of 120,000 people in the United Kingdom compared the risk of COVID-19 (severe enough to be diagnosed in a hospital or emergency room) between workers in eight essential occupations and nonessential workers (Mutambudzi et al. 2021). The highest relative risk was for healthcare professionals, medical support staff, health associate professionals, and social care workers. Healthcare workers had more than a seven-fold increase in COVID-19 risk. A strong point of this study is that it was able to control for a wide range of health and demographic covariates.

Song et al. (2021) used detailed private insurance data from Pennsylvania to compare COVID-19 infection rates for essential and nonessential workers. Essential workers were defined as those employed by firms designated by the governor as life-sustaining businesses and permitted to continue physical operations during a statewide shutdown. Because this definition required physical operation, many of the workers involved were likely to have contact with others. The authors use a difference-in-differences specification to see if essential workers had higher infection rates following the shutdown order. They find that for a sample of policyholders younger than 65, “being an essential worker is associated with a 53% increase in likelihood of being Covid-positive” (p. 11). This increase was most notable in the health and social care sectors, but still present to a smaller extent for essential workers outside those sectors.

Two other studies use death records to assess the occupational mortality risk of COVID-19. A Swedish study found that taxi and bus drivers and service sector workers had a higher risk of dying from COVID-19 than other workers, but the elevated risk disappeared when demographic covariates were included (Billingsley et al. 2020). A California study estimated excess deaths for nine occupational sectors over the period March–November 2020 (Chen et al. 2021). The sectoral highs were in food or agriculture, transportation or logistics, manufacturing, and facilities. Healthcare workers faced only slightly elevated mortality risk. A weakness of this study is that the authors have little information on covariates, as well as imperfect data on occupation.

Based on this evidence, it appears that hazard pay would have made sense during the pandemic. There was indeed a portion of the economy that needed to remain open but that involved a greater risk of illness. At the same time, the evidence suggests that the fraction of essential workers at noticeably higher risk was relatively small. Both Magnusson et al. (2021) and Mutambudzi et al. (2021) identify healthcare workers as being the main workers at greater risk (at least early in the pandemic). Using occupational data from the May 2019 National Occupational Employment and Wage Estimates, healthcare workers (broadly defined) account for about 10 percent of all US workers (US Bureau of Labor Statistics 2019). Expanding the list to include food service workers, taxi drivers, and cleaners, the other noticeably higher risk occupations, adds another 8 percent of US workers.

Compared with this list, the HEROES Act had an excessively broad definition of who should receive hazard pay during the pandemic. The list of eligible occupations or industries in the bill went on for more than four pages and included not just the obvious ones of healthcare and public transportation, but child care, barge operations, longshoremen, laundry work, and restaurant work. The list of eligible industries in the Pennsylvania program was shorter, but still included many for which there is little evidence of higher risk, such as security services, freight trucking, and retail food stores. Both programs did limit the bonuses to frontline workers—that is, essential workers who were required to interact with others. Blau, Koebe, and Meyerhofer (2021) estimate that in March 2020, about 43 percent of all US workers met these criteria.

Both the HEROES Act and the Pennsylvania hazard pay program relied on firms to apply for hazard pay funds from the government. From a social insurance perspective, this feature is problematic. Relying on employer application risks identical workers being treated differently (or worse, workers facing less risk getting premium pay and workers facing more risk not). Whether workers receive the government-provided hazard pay could depend on employer motivation or simply on how quickly employers respond to the call for applications. These outcomes are not socially optimal.

Various alternatives are possible. A minor variation that would reduce the possible randomness would be for the government to invite firms in the relevant industries to say how many workers they have eligible for the hazard pay. The presumption would be that all such identified workers would receive it. Firms would

still have to make some effort for their workers to receive the bonus, but the burden would be less. A more extreme variation would be to require individual workers to apply, as they do for unemployment insurance. This proposal has the great disadvantage of being administratively complex. Not only would the government need to process a very large number of applications, but firms would likely still have to be involved because they are the sensible unit for dispensing the pay.

How Large Should the Hazard Premium Be?

The HEROES Act called for a hazard premium of \$13 per hour for all eligible workers; the Pennsylvania program provided a boost of \$3 per hour. Since risk is related to contact, and contact is time-dependent, it is appropriate for the hazard premium to be per hour worked rather than lump sum. A more difficult and fundamental question is just how much extra pay would be appropriate for high-risk workers.

A crude way to try to answer this question is to blend risk estimates with an estimate of the value of a statistical life. Table 3 presents such a calculation. The Centers for Disease Control and Prevention (2021a) provide an estimate of the number of COVID-19 infections for adults 18 to 64 in the United States over roughly the first year of the pandemic. We combine this with data on the working-age population in 2019 from the US Census Bureau (Rogers and Wilder 2020) to get an overall infection rate (row 1 of Table 3). Magnusson et al. (2021, Figure 3) find that healthcare and public transit workers in Norway had a risk of contracting COVID-19 that was approximately three times greater than other working-age adults (adjusting for age and sex) during the first wave of the pandemic. As discussed above, these workers comprise about 10 percent of employed workers in the United States, or 7 percent of the working-age population in 2019. From this, we can back out estimates of the risk of getting COVID-19 during the first year of the pandemic for both the high-risk group and other working-age adults (row 2). This calculation suggests that high-risk workers had a 59 percentage point higher risk of infection.⁶ We then estimate the infection fatality rate for working-age adults by dividing confirmed working-age deaths (Centers for Disease Control 2021b) by estimated cases (row 3).⁷ The Environmental Protection Agency and other government agencies use \$10 million as their estimate of the value of a statistical life (row 4) (for example, US Environmental Protection Agency 2016).

Armed with these components, we can calculate the cost to high-risk workers of the extra risk of dying they incurred during the pandemic. Multiplying the excess risk times the infection fatality rate times \$10 million yields \$8,617 (row 5

⁶This difference is almost surely somewhat too large. It may reflect the fact that the Magnusson et al. (2021) estimates for relative risk come from early in the pandemic when overall infection rates in Norway were very low. It is worth reiterating, however, that Mutambudzi et al. (2021) also estimate very high risk for healthcare workers.

⁷Our resulting infection fatality rate of 0.146 percent matches quite closely the estimates from O'Driscoll et al. (2020) for the infection fatality rate for working-age adults based on a meta-analysis of studies for a number of countries.

Table 3

Putting a Dollar Value on Increased COVID-19 Risk for High-Risk Workers

1.	Share of working-age population who had COVID-19 during first year of the pandemic: $75,302,292/213,610,414 = 0.35$
2.	Estimated risk of getting COVID-19 based on Magnusson et al. (2021): Other working age population: 29% Healthcare and transit workers: 88% Difference: 59 percentage points
3.	Estimated infection fatality rate for working-age population: $110,143/75,302,292 = 0.00146$
4.	Value of a statistical life: \$10 million
5.	Dollar value of excess risk to healthcare and transit workers: $0.59 \cdot 0.00146 \cdot \$10 \text{ million} = \$8,617 \text{ per year (or } \$4.13/\text{hour)}$
6.	Dollar value of excess risk to all frontline workers based on Magnusson et al.: $0.18 \cdot 0.00146 \cdot \$10 \text{ million} = \$2,704 \text{ per year (or } \$1.30/\text{hour)}$
7.	Dollar value of excess risk to top two risk tiers based on Magnusson et al.: <i>Healthcare and transit workers:</i> $0.59 \cdot 0.00146 \cdot \$10 \text{ million} = \$8,617 \text{ per year (or } \$4.13/\text{hour)}$ <i>Food service workers, taxi drivers, and cleaners:</i> $0.29 \cdot 0.00146 \cdot \$10 \text{ million} = \$4,308 \text{ per year (or } \$2.06/\text{hour)}$

Note: Equations may not hold exactly because of rounding.

of Table 3). The pandemic had been going on for roughly a year as of the time of the data used in the calculation, so this estimate corresponds to the pay boost that would be needed for a year to compensate for the additional risk. This works out to \$4.13 per hour for a typical 2,087-hour work year (US Office of Personnel Management 2021). This is roughly what the Pennsylvania hazard pay program provided, but substantially less than what the HEROES Act called for.

There are various ways that this estimate of the appropriate size of the hazard premium could be too low. Most obviously, it only includes the risk of dying from COVID-19. Studies suggest that at least 10 percent of COVID-19 survivors experience long-term effects (for example, Greenhalgh et al. 2020). Workers might put a substantial dollar value on that possibility as well. Second, hazard pay is motivated by a desire to ensure that high-risk workers remain willing to do socially valuable work. As a result, what matters is not necessarily their actual risk of infection and death, but their perceived risks. If high-risk workers systematically overestimate their risks, the hazard premium might need to be higher. For example, if the perceived infection fatality rate were twice as high as the estimate in Table 3, then the perceived cost of excess risk to high-risk healthcare and transit workers would be over \$8 per hour—getting closer to the hazard premium called for in the HEROES Act. Third, the calculations in Table 3 use confirmed COVID-19 deaths to calculate the

infection fatality rate. Excess deaths (from all causes) are estimated to be approximately 30 percent higher than confirmed COVID-19 deaths in 2020 (Islam et al. 2021). Using this alternative estimate increases the hazard pay needed by 30 percent as well.

Importantly, the calculation above is only for the highest-risk workers. The HEROES Act proposed to provide hazard pay for most frontline workers. The high-contact workers analyzed by Magnusson et al. (2021) account for roughly 31 percent of all workers in the United States in 2019.⁸ For this broader group, the overall weighted average of the added risk of infection (adjusted for age and sex) was a factor of 1.59 (which is very similar to the factor of 1.53 found by Song et al. 2021), which translates into an elevated infection risk of 18 percentage points. Line 6 of Table 3 shows that the resulting dollar value of the extra risk to frontline workers as a group is relatively low—precisely because most face little elevated risk (or in some cases lower risk than non-frontline workers). The calculation suggests that the dollar value of the added risk for this broad group is \$2,704, or roughly \$1.30 per hour.

Given the variation in risk among frontline workers, an appealing possibility would be to have some sort of tiered hazard pay system. While neither the evidence nor administrative capacity is likely to be adequate for a highly variable scale, it seems possible that having a few tiers could result in a program that more closely fits the social insurance ideal. For example, in the Magnusson et al. (2021) study, roughly 10 percent of workers have a threefold increase in risk, and another 8 percent have a twofold increase. As shown in line 7 of Table 3, these numbers imply a dollar value of the extra risk of \$4.13 per hour for the highest-risk workers, and \$2.06 for the second highest risk category.⁹

While these various calculations do not yield a firm number for the size of any hazard premium, they illustrate how policymakers might go about figuring out an appropriate premium in a pandemic. A crucial input, particularly for a tiered hazard pay scheme, is good evidence on the degree of elevated risk for different occupations and industries.

There are obviously other ways policymakers could try to ascertain the appropriate amount of hazard pay. For example, they could gather data on labor shortages in frontline industries. During the pandemic, there was a great deal of anecdotal evidence of labor shortages, particularly for nurses and home healthcare workers (for example, McLernon 2020; Nguyen 2020). Though it would be impractical to try to estimate a full labor supply function in real time, the degree of labor shortage in frontline industries could provide a rough guide for at least the appropriate ranking of the hazard premium across industries.

⁸We calculate this percentage by matching the categories of workers in the Magnusson et al. study with those from the US Bureau of Labor Statistics (2019).

⁹For the increased risk for food service workers, taxi drivers, and cleaners, we use the estimates from the second wave of infection, when the relevant sectors were more likely to be open (see Magnusson et al. 2021, Figure 4).

Should Hazard Pay Phase out at High Incomes?

Whether hazard pay should phase out above some income level depends on the relationship between the utility of income from working and the disutility of working itself for high-wage, high-risk workers. If the two were reasonably close before the pandemic, the rise in the disutility of work caused by the pandemic could cause many of these workers to stop working (even if they were not eligible for unemployment insurance). Because such workers are very valuable to society, in this case hazard pay should continue even at high incomes. On the other hand, if the utility provided by the income from working was substantially above the disutility of working for high-income workers in the high-risk sector before the pandemic, even a substantial rise in their disutility of work would not make them choose unemployment. In this case, it would make sense for hazard pay to phase out.

There are good reasons for thinking that the scenario where high-income workers in the risky sector choose to continue working despite increased risk is the more plausible one. Workers such as doctors and dentists have invested years in training and building a practice. The present value of the income loss they would face from leaving their jobs, even temporarily, is very large. More generally, it is likely that high human capital and high match quality create a substantial wedge between what high-skilled, high-risk workers earn and the disutility of work. Finally, there is also evidence that nonpecuniary job benefits are higher for well-educated workers (see Duncan 1977). This too creates a wedge between the wage and the disutility of labor that would encourage high-income, high-risk workers to continue working even if the disutility of work rises. Thus, it would likely be appropriate for hazard pay to phase out at high incomes.

Interestingly, the HEROES Act did not phase out hazard pay entirely for high-wage workers, though it did limit it. The act capped the amount of premium pay at \$10,000 for eligible workers earning less than \$200,000, and \$5000 for eligible workers earning more than \$200,000. The Pennsylvania program, on the other hand, only applied to workers earning less than \$20 per hour (excluding benefits and overtime).

Hazard Pay as Demand Stimulus

The primary motivation for hazard pay during a pandemic from a social insurance perspective is to make sure essential workers in hazardous sectors are not worse off than workers who become unemployed because their sectors shut down. However, because there may be inadequate aggregate demand during a pandemic, an important practical consideration is whether hazard pay is likely to be useful as demand stimulus. From a logical standpoint, hazard pay wouldn't seem to have a particularly large bang for the buck. By definition, the workers receiving it remain employed. As a result, one might think that the recipients have a lower marginal propensity to consume than the unemployed workers discussed in the previous section.

One factor that militates against this presumption is that many frontline workers are relatively low-income. Blau, Koebe, and Meyerhofer (2021, p. 172) estimate

that frontline workers had lower average hourly wages (\$22.76 versus \$27.05 for all workers). This is even true for healthcare workers, where the case for hazard pay is strongest. While healthcare practitioners and technicians earn more than \$40 per hour, Blau, Koebe, and Meyerhofer (2021, p. 172) find that healthcare support workers earn less than \$20 per hour. The other five lowest-wage frontline occupations the authors identify are food preparation and serving; building and grounds cleaning and maintenance; personal care and service occupations; farming, fishing, and forestry; and transportation and material moving. Many of these are occupations that would be next in line after healthcare to receive hazard pay based on their relative COVID-19 risk.

Other studies reach similar conclusions about the low-wage nature of much elevated-risk work. Kinder, Stalder, and Du (2020) calculate that “as of 2018, nearly half (47%) of all frontline essential workers earned less than a living wage” (which they identify as \$16.14 in 2018). For example, personal care aides had a median wage of \$11.55, and janitors and cleaners had a median wage of \$12.55. Kearney and Muñana (2020), using the Kaiser Family Foundation Tracking Poll, find that essential workers working outside the home were more likely to earn less than \$40,000 than other currently employed workers (31 percent versus 19 percent). Forty-nine percent of them said they would struggle to pay a \$500 unexpected medical bill, while only 31 percent of other workers said they would.

That high-risk workers are disproportionately low-wage and would struggle to pay bills suggests they are likely to have a higher marginal propensity to consume than a typical worker. A higher MPC should translate into a higher fiscal multiplier. Thus, hazard pay could pack a larger stimulatory impact than broader types of stimulus, such as general tax cuts or widely available one-time payments. This extra stimulatory impact could be enhanced by limiting hazard pay to low-wage workers, as in the Pennsylvania program.

Conclusion

In many ways, our social insurance perspective on fiscal policy harkens back to an older literature. Rather than focusing on aggregate demand management, we emphasize the more traditional role of government in providing insurance against life’s vicissitudes. This perspective is particularly appropriate for fiscal policy during a pandemic, when output needs to remain low in some sectors for health reasons, and aggregate demand stimulus cannot flow to many affected workers because their sectors are shut.

Our framework shows that thinking in terms of social insurance leads naturally to directing government aid to those directly harmed by the pandemic—particularly the unemployed and those who work in essential jobs with a high risk of infection. Our more practical analysis of unemployment insurance and hazard pay shows how a social insurance perspective can provide guidance on who should receive these benefits, how large they should be, and how long they should last. An analysis of the

policies actually taken or proposed for unemployment insurance and hazard pay in the United States during the pandemic shows that many of the actions follow what a social insurance perspective would recommend, but a number did not.

An obvious question about the social insurance perspective on fiscal policy is whether it's likely to have usefulness beyond the COVID-19 pandemic. We believe it does. First, it is all too likely that there will be future pandemics. The implications of the social insurance perspective we have discussed could yield much more successful fiscal policy in any future public health crises.

Second, it's possible that the United States and other countries could face more regionally concentrated recessions in the future that share important similarities to a pandemic. For example, parts of the American West and Australia have experienced prolonged economic disruption due to drought-induced fires. As with a pandemic, such natural-disaster-fueled downturns aren't easily remedied with broad aggregate demand stimulus. They are also likely to involve high risks to certain essential workers. The lessons derived from a social insurance perspective for pandemics about targeting relief to those most affected and paying a hazard premium to essential high-risk workers are likely to carry over to these downturns as well.

Finally, while the unique features of the pandemic have made a social insurance perspective vital in the current situation, it may have value in more ordinary downturns. Most recessions involve highly unequal impacts on different types of workers, and general stimulus often takes a long time to help some workers regain employment. A fiscal response that focuses on both aggregate demand management and social insurance might prove more effective in dealing with future recessions regardless of their cause.

■ *We are grateful to Jacob Weber for research assistance and to Peter Ganong and the editors for helpful comments.*

References

- Baker, Scott R., R.A. Farrokhnia, Steffen Meyer, Michaela Pagel, and Constantine Yannelis.** 2021. "Income, Liquidity, and the Consumption Response to the 2020 Economic Stimulus Payments." <https://drive.google.com/file/d/19KaPY-xV27feRZF74EBby3DZ3RW9K3Bp/view>.
- Billingsley, Sunnee, Maria Brandén, Siddhartha Aradhya, Sven Drefahl, Gunnar Andersson, and Eleonora Mussino.** 2020. *Deaths in the Frontline: Occupation-Specific COVID-19 Mortality Risk in Sweden*. Stockholm: Stockholm University Demography Unit.
- Bitler, Marianne P., Hilary W. Hoynes, and Diane Whitmore Schanzenbach.** 2020. "The Social Safety Net in the Wake of COVID-19." *Brookings Papers on Economic Activity* (Summer): 119–45.

- Blau, Francine D., Josefine Koebe, and Pamela A. Meyerhofer.** 2021. "Who Are the Essential and Front-line Workers?" *Business Economics* 56: 168–78.
- Cajner, Tomaz, Andrew Figura, Brendan M. Price, David Ratner, and Alison Weingarden.** 2020. "Reconciling Unemployment Claims with Job Losses in the First Months of the COVID-19 Crisis." Federal Reserve Board, Finance and Economics Discussion Series 2020–055.
- Centers for Disease Control and Prevention.** 2021a. "Estimated Disease Burden of COVID-19." <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html> (accessed May 1, 2021).
- Centers for Disease Control and Prevention.** 2021b. "Weekly Updates by Select Demographic and Geographic Characteristics." https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm (accessed May 1, 2021).
- Chen, Yea-Hung, Maria Glymour, Alicia Riley, John Balmes, Kate Duchowny, Robert Harrison, Elliott Matthay, Kirsten Bibbins-Domingo.** 2021. "Excess Mortality Associated with the COVID-19 Pandemic among Californians 18–65 Years of Age, by Occupational Sector and Occupation: March through November 2020." *PLoS ONE* 16 (6).
- Chodorow-Reich, Gabriel, and John Coglianesi.** 2019. "Unemployment Insurance and Macroeconomic Stabilization." In *Recession Ready: Fiscal Policies to Stabilize the American Economy*, edited by Heather Boushey, Ryan Nunn, and Jay Shambaugh, 153–79. Washington, D.C.: Brookings Institution.
- Duncan, Greg J.** 1977. "Earnings Functions and Nonpecuniary Benefits." *Journal of Human Resources* 11 (4): 462–83.
- Farrell, Diana, Peter Ganong, Fiona Greig, Max Liebeskind, Pascal Noel, and Joseph Vavra.** 2020. "Consumption Effects of Unemployment Insurance during the COVID-19 Pandemic." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3654274.
- Ganong, Peter, Fiona Greig, Max Liebeskind, Pascal Noel, Daniel M. Sullivan, and Joseph Vavra.** 2021. "Spending and Job Search Impacts of Expanded Unemployment Benefits: Evidence from Administrative Micro Data." Becker Friedman Institute Working Paper 2021-19.
- Ganong, Peter, and Pascal Noel.** 2019. "Consumer Spending during Unemployment: Positive and Normative Implications." *American Economic Review* 109 (7): 2383–2424.
- Ganong, Peter, Pascal Noel, and Joseph Vavra.** 2020. "US Unemployment Insurance Replacement Rates during the Pandemic." *Journal of Public Economics* 191.
- Greenhalgh, Trisha, Matthew Knight, Christin A'Court, Maria Buxton, and Laiba Husain.** 2020. "Management of Post-Acute COVID-19 in Primary Care." *British Medical Journal* 370: m3726.
- Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning.** 2020. "Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?" https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/6/2265/files/2019/04/macro_covid_GLSW_July_2020.pdf.
- Islam, Nazrul, Vladimir M. Shkolnikov, Rolando J. Acosta, Ilya Klimkin, Ichiro Kawachi, Rafael A. Irizarry, Gianfranco Alicandro et al.** 2021. "Excess Deaths Associated with COVID-19 Pandemic in 2020: Age and Sex Disaggregated Time Series Analysis in 29 High Income Countries." *British Medical Journal* 373: n1137.
- Karger, Ezra, and Aastha Rajan.** 2021. "Heterogeneity in the Marginal Propensity to Consume: Evidence from COVID-19 Stimulus Payments." Federal Reserve Bank of Chicago Working Paper 2020–15.
- Kearney, Audrey, and Cailey Muñana.** 2020. "Taking Stock of Essential Workers." *Kaiser Family Foundation*, May 1. <https://www.kff.org/policy-watch/taking-stock-of-essential-workers/>.
- Kinder, Molly, Laura Stater, and Julia Du.** 2020. *The COVID-19 Hazard Continues, but the Hazard Pay Does Not: Why America's Essential Workers Need a Raise*. Washington, DC: Brookings.
- Magnusson, Karin, Karin Nygård, Fredrik Methi, Line Vold, and Kjetil Telle.** 2021. "Occupational Risk of COVID-19 in the 1st vs 2nd Wave of Infection." <https://doi.org/10.1101/2020.10.29.20220426>.
- McLernon, Lianna Matt.** 2020. "COVID-Related Nursing Shortages Hit Hospitals Nationwide." *Center for Infectious Disease and Research Policy News*, November 30. <https://www.cidrap.umn.edu/news-perspective/2020/11/covid-related-nursing-shortages-hit-hospitals-nationwide>.
- Milne, Alistair.** 2020. "A Critical COVID-19 Economic Policy Tool: Retrospective Insurance." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3558667.
- Mutambudzi, Miriam, Claire Niedzwiedz, Ewan Beaton Macdonald, Alastair Leyland, Frances Mair, Jana Anderson, Carlos Celis-Morales et al.** 2021. "Occupation and Risk of Severe COVID-19: Prospective Cohort Study of 120 075 UK Biobank Participants." *Occupational and Environmental Medicine, British Medical Journal* 78 (5): 307–14.
- Nguyen, Daisy.** 2020. "US Hospitals Facing Worrisome Shortages of Nurses, Doctors." *Associated*

- Press*, December 22. <https://www.usnews.com/news/business/articles/2020-12-22/california-desperately-searches-for-more-nurses-and-doctors>.
- O’Driscoll, Megan, Gabriel Ribeiro Dos Santos, Lin Wang, Derek A.T. Cummings, Andrew S. Azman, Juliette Paireau, Arnaud Fontanet, Simon Cauchemez, and Henrik Salje.** 2020. “Age-Specific Mortality and Immunity Patterns of SARS-CoV-2.” *Nature* 590: 140–45.
- Pennsylvania Department of Community and Economic Development.** 2020. “COVID-19 PA Hazard Pay Grant: Program Guidelines.” <https://dced.pa.gov/download/covid-19-pa-hazard-pay-grant-guidelines-2020/?wpdmdl=95560>.
- Rogers, Luke, and Kristie Wilder.** 2020. “Working-Age Population Not Keeping Pace with Growth in Older Americans.” <https://www.census.gov/library/stories/2020/06/working-age-population-not-keeping-pace-with-growth-in-older-americans.html>.
- Rowe, Nick.** 2020. “Relative Supply Shocks, Unobtainium, Walras’ Law, and the Coronavirus.” *Worthwhile Canadian Initiative: A Mainly Canadian Economics Blog*. https://worthwhile.typepad.com/worthwhile_canadian_initi/2020/03/relative-supply-shocks-unobtainium-walras-law-and-the-coronavirus.html.
- Saez, Emmanuel, and Gabriel Zucman.** 2020. “Keeping Businesses Alive: The Government Will Pay.” <https://econfp.org/wp-content/uploads/2020/03/20.Keeping-Businesses-Alive.pdf>.
- Song, Hummy, Ryan M. McKenna, Angela T. Chen, Guy David, and Aaron Smith-McLallen.** 2021. “The Impact of the Non-Essential Business Closure Policy on COVID-19 Infection Rates.” *International Journal of Health Economics and Management* 21 (4): 387–426.
- US Bureau of Labor Statistics.** 2019. “May 2019 National Occupational Employment and Wage Estimates, United States.” https://www.bls.gov/oes/2019/may/oes_nat.htm (accessed September 19, 2021).
- US Congress.** 2020. “H.R. 6800.” 116th Congress, 2d Session. <https://www.congress.gov/bill/116th-congress/house-bill/6800/text>.
- US Environmental Protection Agency.** 2016. “Valuing Mortality Risk Reductions for Policy: A Meta-Analytic Approach.” [https://yosemite.epa.gov/sab/sabproduct.nsf/0/0CA9E925C9A702F285257F380050C842/\\$File/VSL%20white%20paper_final_020516.pdf](https://yosemite.epa.gov/sab/sabproduct.nsf/0/0CA9E925C9A702F285257F380050C842/$File/VSL%20white%20paper_final_020516.pdf).
- US Office of Personnel Management.** 2021. “Fact Sheet: Computing Hourly Rates of Pay Using the 2,087-Hour Divisor.” <https://www.opm.gov/policy-data-oversight/pay-leave/pay-administration/fact-sheets/computing-hourly-rates-of-pay-using-the-2087-hour-divisor/>.
- Weinzierl, Matthew.** 2014. “The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice.” *Journal of Public Economics* 118: 128–42.
- Woodford, Michael.** 2020. “Effective Demand Failures and the Limits of Monetary Stabilization.” NBER Working Paper 27768.

Should We Insure Workers or Jobs During Recessions?

Giulia Giupponi, Camille Landais, and Alice Lapeyre

In the wake of the COVID crisis, labor market policy responses on both sides of the Atlantic have been immediate, absolutely unprecedented in scope—and also diametrically opposed in nature. To put it simply, the focus of the US labor market policy response was on insuring the income of workers against the cost of job losses. This was done by aggressively increasing the generosity of unemployment insurance. In Europe, the emphasis was on preserving the relationship between workers and firms, which translated into generous subsidies for hours reductions and temporary layoffs through short-time work or related schemes.

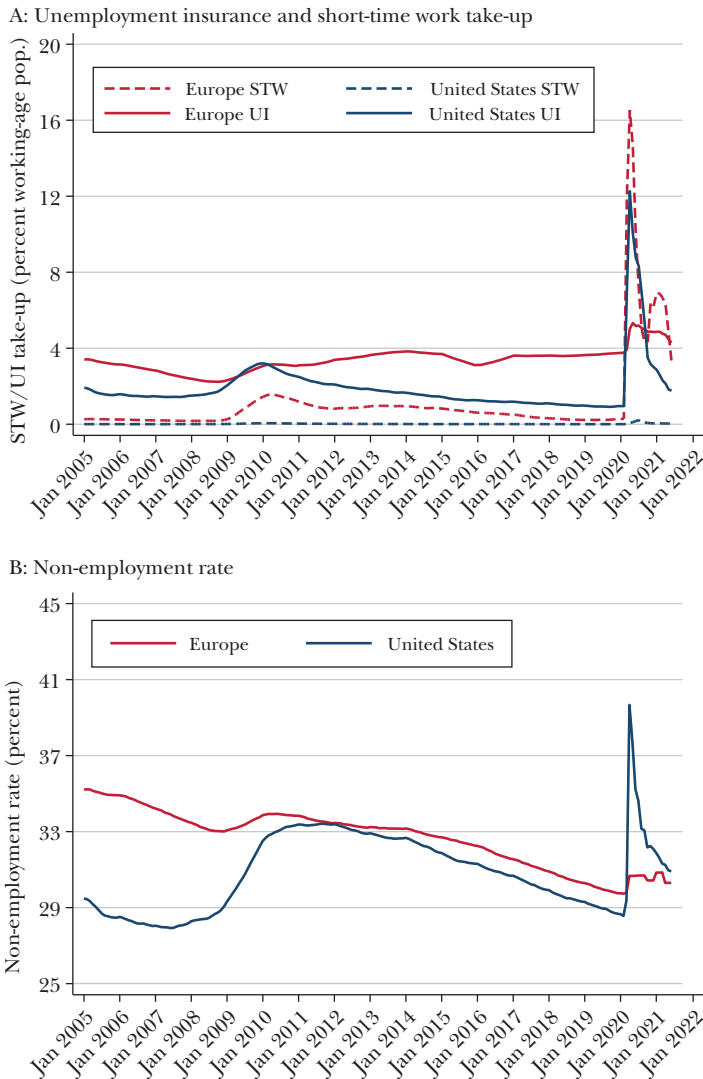
Panel A of Figure 1 gives a visual representation of these polar strategies. In the United States, the fraction of the working-age population on unemployment insurance benefits surged from about 2 to 12 percent in April 2020, and, although it declined very quickly after that, at the end of 2020 it was still higher than at the peak of the Great Recession of 2007–09. If we look at the weighted sum of the four largest European economies—Germany, the United Kingdom, France, and Italy—the increase in the number of unemployment insurance recipients was very limited, but take-up of short-time work skyrocketed, with more than 16 percent of the working-age population enrolled in this type of scheme in April 2020. There was no such increase in short-time work take-up in the US economy, although about 25 US states have operational work-sharing schemes similar to short-time work.

■ *Giulia Giupponi is Assistant Professor of Economics, Bocconi University, Milan, Italy. Camille Landais is Professor of Economics, London School of Economics, London, United Kingdom. Alice Lapeyre is a PhD candidate in Economics, CREST-ENSAE (Center for Research in Economics and Statistics, École Nationale de la Statistique et de l'Administration Économique), Paris, France. Their email addresses are giulia.giupponi@unibocconi.it, c.landais@lse.ac.uk, alice.lapeyre@ensae.fr.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.29>.

Figure 1

Labor Market Policy Responses to Recessions and Non-Employment Rates in the United States and Europe



Note: Panel A reports the evolution of short-time work (dashed lines) and unemployment insurance (solid lines) take-up in Europe (red lines) and the United States (blue lines), each computed as the ratio of the number of individuals in the program in a given month, as a percent of the quarterly working age population. The series for Europe is a weighted average of the series for Germany, France, Italy, and the United Kingdom, weighted by the working-age population. Panel B reports the evolution of the non-employment rate—that is, one minus the employment rate (and thus including both the unemployed and those out of the labor force). In both panels, the plotted series are moving averages of the raw series over the period up to June 2021. The moving average is based on twelve lagged terms, one forward term, and uniform weights. Data on employment come from OECD. Data on short-time work and unemployment insurance take-up come from the OECD and national statistics. See Online Appendix C for details on data sources and the construction of short-time work/unemployment insurance take-up.

Some consequences of these opposite labor market strategies on non-employment rates are laid bare in Panel B. While the US economy experienced a spike in non-employment, and continued to see high rates of non-employment in late 2020, employment rates did not bulge in Europe despite the severity of the shock. Interestingly, the much larger cyclical volatility of the US labor market relative to that of European countries was already visible in past recessions, during which Europe already experimented, although to a much lower degree, with short-time work usage. Do US policymakers get it right by focusing their labor market policy response to recessions on insuring workers through unemployment insurance? Or should they use more short-time work and focus more on preserving jobs, as in Europe?

Addressing these questions is complicated by the remarkably small attention devoted to short-time work relative to the sprawling literature on unemployment insurance—an imbalance we hope to remedy.

Short-Time Work and Unemployment Insurance

Some Institutional Features

While most people are familiar with unemployment insurance policies, short-time work schemes are not as well-known. How do they work in practice? What countries use them and how long have they been in place? How do they compare with unemployment insurance in terms of generosity, coverage, or eligibility? Let us start by clarifying a bit the institutional background.

Short-time work—also known as short-time compensation, work sharing, or shared-work programs—is a subsidy for temporary reductions in the number of hours worked in firms experiencing temporary drops in demand or production. Short-time work programs allow employers facing temporary shocks to reduce their employees' hours instead of laying them off. The program provides a subsidy to employees put on reduced hours (that is, put on short-time work) equivalent to a fraction of their lost earnings. Short-time work cushions the adverse effect of reductions in business activity on both firms and workers, averting the risk of layoffs and insuring workers against the cost of drops in hours worked. Unemployment insurance programs, instead, provide a temporary subsidy to laid-off workers who lost their job through no fault of their own. Hence, while both programs provide insurance against labor market shocks, short-time work programs insure job matches against temporary hours reductions, while unemployment insurance insures workers against job loss. We now provide a general description of the features of both programs.

Several European countries and US states have short-time work schemes in place.¹ Whilst different schemes are characterized by different requirements and generosity, they all share a common feature in that the application process must be initiated by

¹At the time of writing, 25 states are operating active programs in the United States: Arizona, Arkansas, California, Colorado, Connecticut, Florida, Iowa, Kansas, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, New Hampshire, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, Texas, Washington, and Wisconsin. Vermont ceased to operate its short-time work program on July 1, 2020.

the employer, who must submit an application to the relevant administrative agency. If successful, the application grants a number of subsidized hours of short-time work that can be used by the firm. Workers' subsidies are computed as a percentage of the earnings lost due to hours not worked, typically up to a cap. The disbursement of the subsidy is usually advanced to the worker by the firm, who is subsequently reimbursed by a social insurance administration via lower contribution payments. In exceptional cases, the social insurance agency pays the subsidy directly to the worker. Short-time work schemes are usually funded through a combination of social insurance contributions—paid by eligible firms and workers—and an experience rating component, paid only by those firms and workers that benefit from the program.

Short-time work regulations can include *work-sharing requirements*, specifying a minimum and/or maximum hours reduction, a minimum number of employees or share of the workforce to be involved in the program, and how hours reductions should be distributed across the workforce. In the majority of countries, there is no maximum hours reduction per worker, meaning that short-time work can cover both partial and full hours reductions (that is, down to zero hours worked). In the United States, instead, work weeks must be reduced by at least 10 percent and no more than 60 percent to benefit from the program.

Eligibility requirements set the conditions under which employers or employees can take part in the scheme. For employers, the main requirement is proof of economic need, such as a reduction in business activity above a certain threshold. In some countries, access to the program is restricted to firms operating in certain sectors of the economy or with size above given thresholds. An agreement with union representatives can also be required. To be eligible for short-time work, workers are typically required to have been employed and contributing to social insurance for a minimum amount of time. Some short-time work schemes do not cover workers on temporary contracts.

Conditionality requirements for employers can include the prohibition of dismissals and the development of a recovery plan, while for workers, either training or job search requirements could be in place.²

Unemployment insurance provides temporary financial assistance to eligible unemployed workers who become unemployed through no fault of their own. To receive unemployment insurance, laid-off workers must file a claim with the relevant administrative agency. To be eligible, a worker is typically required to have worked for a minimum amount of time, and in some cases to have earnings above a certain threshold. Workers on fixed-term contracts are usually not covered by unemployment insurance once their contract expires. If eligible, the unemployed worker receives an unemployment benefit in cash for a given amount of time. Benefits are defined as a fraction of recent earnings, usually up to a cap, with a replacement rate that is in most cases lower than that of short-time work at zero hours (and the more so at partial hours). The disbursement of unemployment benefits is made directly to the worker by the social insurance agency, typically ensuring swift payments. To retain their

²For a comprehensive illustration of short-time work schemes in OECD countries, see Hijzen and Venn (2011).

benefits, unemployed workers may need to fulfill job-search and availability-to-work requirements—features aimed at ensuring that recipients do not become inactive.

Whilst in their purest form, as we have described them so far, short-time work and unemployment insurance are polar schemes insuring jobs on the one hand and workers on the other, in practice there exists a continuum of more nuanced labor market policies between those two extremes. Taking the United States as an example, workers can qualify for unemployment insurance both when on permanent and temporary layoffs. Temporary unemployment—also called “recall” unemployment—is a situation in which an employer lays off a worker or a group of workers but plans to rehire them by a given date. In this circumstance, the worker is eligible for unemployment benefits and job search requirements are usually waived. Employers also have the option to put full-time employees who work less than full time during a pay period due to lack of work on partial unemployment insurance. In this case, workers are eligible for partial unemployment benefits, provided that they do not earn more than a maximum amount of labor income per week. Apart from this, eligibility conditions for partial unemployment insurance are usually identical to those for full unemployment insurance.

We can therefore think of there being a spectrum of policies offering different types of flexibility to employers: from short-time work for partial hours reductions offering only intensive margin flexibility, to short-time work at zero hours and partial unemployment offering both intensive and extensive margin flexibility, to recall unemployment offering only flexibility at the extensive margin. Besides differences in flexibility, the programs also insure different shocks and, consequently, are characterized by different types of moral hazard responses. On the one hand, programs that ensure intensive-margin adjustments tend to insure job matches: they trigger moral hazard responses on the employer side, because employers may have incentives to rely excessively on subsidized hours reductions. This possibility is especially relevant in contexts—like European countries—in which short-time work is only mildly experience-rated. On the other hand, unemployment insurance schemes insure workers rather than jobs and trigger moral hazard responses in the form of lower job search effort exerted by the unemployed.

It is worth noting that, whilst at first sight short-time work and partial or recall unemployment might seem to serve the same insurance needs, they are fundamentally different in how they affect employers’ commitment to retain workers. The possibility to make take-up conditional on the prohibition of dismissal is a key advantage of short-time work programs compared to partial or recall unemployment. For these, it is hard to envisage a mechanism whereby firms are held to their commitment to retain or recall workers, since such a commitment can neither be monitored nor enforced.

Short-Time Work and Unemployment Insurance During the COVID Crisis

At the onset of the COVID crisis, the United States responded to the sudden labor market freeze and historical surge in layoffs by aggressively extending the generosity of unemployment insurance. In particular, the Coronavirus Aid, Relief, and Economic Security (CARES) Act signed into law in March 2020 granted i) additional payments

to everyone who qualified for unemployment benefits; ii) an extension to individuals who would have otherwise exhausted their benefits; and iii) eligibility to self-employed and gig workers. Specifically, the CARES Act authorized Federal Pandemic Unemployment Compensation, in which unemployment benefits were increased by \$600 a week from March to July 2020. Moreover, the CARES Act was complemented by two additional stimulus packages in 2021—the Consolidated Appropriations Act (January) and the American Rescue Plan (March)—both of which extended the unemployment insurance measures put in place by the CARES Act.

European countries, on the contrary, responded through short-time work or related schemes. In April 2020, the European Union announced that it would provide financial assistance for up to €100 billion to EU countries to develop or expand short-time work schemes.³ The majority of OECD countries had a short-time work program in place prior to the COVID crisis: *Activité Partielle* in France, *Kurzarbeit* in Germany, and *Cassa Integrazione Guadagni* in Italy. Several of those who did not have a scheme in place introduced it (as in Hungary and the United Kingdom), and most of those with existing short-time work schemes implemented measures to ensure rapid access to and wide take-up of the program. Combinations of such measures have been necessary for short-time work schemes to work swiftly and effectively (Giupponi and Landais 2020a).⁴

Insurance versus Moral Hazard

Let us imagine that, in the midst of recession, we decide to increase the generosity of social insurance by \$1. Should this \$1 be put into more generous unemployment insurance or into more generous short-time work?

To approach the choice between unemployment insurance and short-time work, our starting point is the standard trade-off that both policies have to solve between providing insurance to workers against labor market shocks and distorting the behaviors of firms and workers (Baily 1978; Chetty 2008). Indeed, the goal of both policies is to provide insurance against labor market shocks. For unemployment insurance, the shock is the cost of being unemployed. For short-time work, it is the cost of having to reduce working hours when a firm is hit by a negative shock.

Providing such insurance is likely to be socially desirable, as it transfers money to individuals who have lower income and consumption. But transferring \$1 to these individuals will cost more than \$1 because these policies also tend to distort behaviors, a problem often called moral hazard. Individuals will search less actively for a new job when they have unemployment insurance and will cash benefits for longer as a result. Firms may reduce hours more than otherwise necessary if their

³For more details on the Support to mitigate Unemployment Risks in an Emergency (SURE) program, see https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/financial-assistance-eu/funding-mechanisms-and-facilities/sure_en.

⁴Online Appendix A provides an overview of measures that have been put in place during the COVID crisis to facilitate access to short-time work. See also Scarpetta et al. (2020) for an overview of short-time work schemes during COVID in OECD countries.

workers receive short-time work benefits. From the perspective of the government, these changes in behavior increase the cost of providing such insurance benefits. We say that moral hazard creates a fiscal externality on the government.

What do we know about the respective moral hazard costs of unemployment insurance and short-time work schemes? And which program is better at providing insurance?

Willingness to Pay for Unemployment Insurance versus Short-Time Work Benefits

The value of any form of social insurance against labor market shocks depends on how workers value insurance (that is, the extent of their risk aversion) as well on whether workers have access to alternative means of consumption smoothing (self-insurance). Estimating this value poses challenges. Given that both unemployment insurance and short-time work are mandated by the government, one cannot simply look at a market price to measure workers' willingness-to-pay for insurance against labor market shocks.

In general, the research literature has devoted relatively little attention to this problem. The literature has mostly focused on measuring consumption dynamics around labor market shocks like job loss (for example, Gruber 1997). This approach usually finds significant but small consumption responses, which in turn translate into a moderate value of social insurance. Recent research using alternative revealed-preference methods (for example, Hendren 2017; Landais and Spinnewijn 2021) suggests instead that the value of insurance against unemployment shocks is much larger than previously thought, but is also strongly heterogeneous across individuals.

For present purposes, the key question is how the insurance value of short-time work compares to that of unemployment insurance. The evidence on this point is limited. However, two elements indicate that the value of unemployment insurance may be somewhat larger than the value of short-time work.

First, recipients of short-time work and recipients of unemployment insurance are often quite distinct populations. For example, Germany is a country where both generous unemployment insurance and short-time work are available. We exploit newly collected data starting in May 2020 from the High-frequency Online Personal Panel Survey (HOPP), a longitudinal survey launched by the German Institute for Employment Research (Haas et al. 2021). It shows unambiguously that during the COVID crisis, short-time work tended to protect mostly insiders, individuals with higher incomes and better self-insurance options. Unemployment insurance, to the contrary, was mostly protecting the outsiders of the labor market, like younger individuals at the beginning of their career, or individuals with lower education and fewer means to smooth household consumption (such as the presence of a working partner).⁵ As noted by Cahuc and Carcillo (2011), because short-time work tends to protect insiders, it is perhaps not surprising that it tends to be more prevalent in countries with strong employment protection regulations.

Second, the value of insurance is a direct function not only of the availability of self-insurance options, but also of the size of the consumption (or income) shock

⁵For these results and additional details on the HOPP survey, see online Appendix B.

experienced upon transitioning to the program. As shown in Figure 2, short-time work clearly insures smaller shocks. The figure builds on administrative data from Germany (Tilly and Niedermayer 2016) and Italy during the Great Recession of 2007–09 (Giupponi and Landais 2020b), and compares, using an event study design, the evolution of total earnings plus transfers around the onset of an unemployment spell and a spell of short-time work. In both panels, we see that the drop in earnings and transfers is much more severe and persistent for the unemployed than for workers on short-time work. But we also see an interesting difference. In Germany, the earnings of workers who experienced a short-time work spell had fully recovered after three years (Panel A). In Italy, to the contrary, they were still 30 percent lower than the year before entering short-time work, and they were converging to the level of earnings of workers having experienced an unemployment shock instead (Panel B). The main explanation for this discrepancy is that the Italian recession was much more protracted, and the shock to firms was therefore much more persistent. This, in turn, reminds us that short-time work tends to insure against temporary shocks, but is less effective at insuring against persistent or permanent shocks: if the shock persists, a firm will not hold onto its workers and will eventually lay them off.

The Relative Moral Hazard Costs of Unemployment Insurance and Short-Time Work

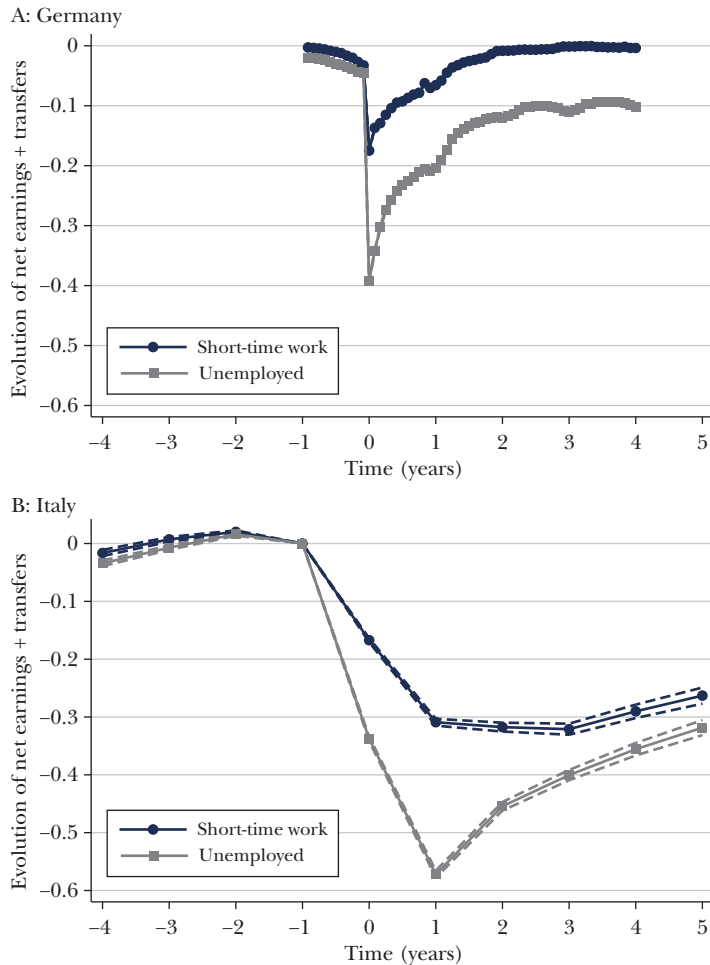
The literature on the moral hazard costs of unemployment insurance is abundant (for example, see Schmieder, von Wachter, and Bender 2016). A main conclusion is that the duration of unemployment spells is strongly responsive to the generosity of unemployment insurance. A smaller literature also investigates the impact of unemployment insurance generosity on the probability of entering an unemployment insurance program, and finds moderate responses. In general, there is less scope for moral hazard along the extensive margin of becoming unemployed (as opposed to the intensive margin of being unemployed for a longer time), as layoffs are well-defined and well-monitored events, and those who quit are quite restricted in their ability to access unemployment insurance in many countries.⁶ Overall, the consensus is that the fiscal externality of increasing the generosity of unemployment insurance is relatively large: the cost to the government of an additional \$1 of unemployment insurance ranges from \$1.50 to \$2.50.

Evidence on the moral hazard costs of short-time work is much more limited, but two elements suggest that these costs might be significant. First, while access to short-time work is generally made conditional on firms experiencing economic or financial distress, the definition of distress is not always very precise and can prove hard to enforce, leaving some room for manipulation. Second, short-time work subsidizes hours *reductions*, which requires an effective monitoring of hours actually worked by employees, a notoriously difficult task for government administrators. As a result, the massive extension across Europe during the COVID crisis of access to

⁶Recent papers hint towards layoff date being responsive to discontinuities in eligibility or generosity of unemployment insurance, one potential explanation being workers and firms bargaining over the timing of the layoff: for example, see Khoury (2021) for evidence on France, and Albanese, Picchio, and Ghirelli (2020) for evidence on Italy.

Figure 2

Evolution of Earnings and Transfers Around the Events of Job Loss and Short-Time Work During the Great Recession



Note: The figure reports the evolution of earnings and transfers around job loss (in grey) or around the start of a short-time work spell (in blue). It shows that job loss is associated with a much larger and much more persistent drop in resources than short-time work, implying that the marginal insurance value is likely greater for unemployment insurance than for short-time work. Panel A reproduces estimates from Tilly and Niedermayer (2016) which uses German administrative data from the Institute for Employment Research (IAB). It corresponds to a weighted average of the effect of short-time work and unemployment insurance on income by tenure category using as weights their share in the population. Panel B reproduces estimates from Giupponi and Landais (2020b) and uses administrative data from INPS on the universe of employer-employee matches and social security payments in the private sector in Italy.

short-time work for small businesses, where the difficulty of monitoring hours can be even more acute, has generated fears of a surge in moral hazard.

The existing evidence on the moral hazard costs of short-time work comes almost exclusively from the Great Recession, but it suggests, interestingly, that these

costs are smaller than anticipated. In the context of Italy, for instance, Giupponi and Landais (2020b) identify behavioral responses to short-time work using variation in eligibility rules across firms, and find that for every €1 transferred to a worker on short-time work during the Great Recession, the total cost to the government implied by behavioral responses was around €1.4. In the context of Switzerland, Kopp and Siegenthaler (2021) compared firms who were successful to firms who were unsuccessful in their short-time work application during the Great Recession and find a negative mark-up: in other words, short-time work paid for itself.⁷

What can explain these relatively small (or even negative) fiscal externalities of short-time work, in contrast with the relatively large moral hazard cost of unemployment insurance? First, it seems that, at least during the Great Recession, there was not much manipulation in the reporting of hours worked.⁸ Second, it appears that the probability of an individual worker being put on short-time work did not respond significantly to the generosity of short-time work subsidies, at least not during the Great Recession. Figure 3 illustrates this point using a large discontinuity in the short-time work subsidy amount available to workers in Italy at a particular wage threshold. Panel A shows, using Italian administrative data, that the average short-time work subsidy increases by 12 percent at the wage threshold. Yet, there is no sign of discontinuity at the threshold in the probability that a worker is put on short-time work, conditional on being employed by a firm using it (Panel C), nor in the share of potential working hours spent on short-time work conditional on being on reduced hours (Panel E).⁹

Figure 3 also shows evidence of large moral hazard responses to unemployment insurance benefits in the same Italian context (Scrutinio 2018). In Italy, potential unemployment benefit duration is entirely based on age at layoff: workers fired before turning 50 are eligible for eight months of unemployment benefits, while workers fired after turning 50 can receive up to twelve months of benefits (Panel B).¹⁰ Panel D plots the density of layoffs by age in months and shows a sharp response to the increase in benefit duration at age 50, suggesting that workers manipulate their layoff date to obtain more generous benefits. Panel F shows that the average number of weeks on unemployment insurance—that is, the intensity of utilization—increases sharply from 23 to 31 weeks at the threshold (based on evidence from Scrutinio 2018).

⁷Evidence from Germany shows that, during the Great Recession, the utilization of short-time work was concentrated among large firms hit by the trade collapse (Boeri et al. 2011). The extension of short-time work schemes to smaller firms during the COVID crisis, as well as the much larger scale of utilization of the program, may limit the applicability of evidence on moral hazard from the Great Recession to the COVID one.

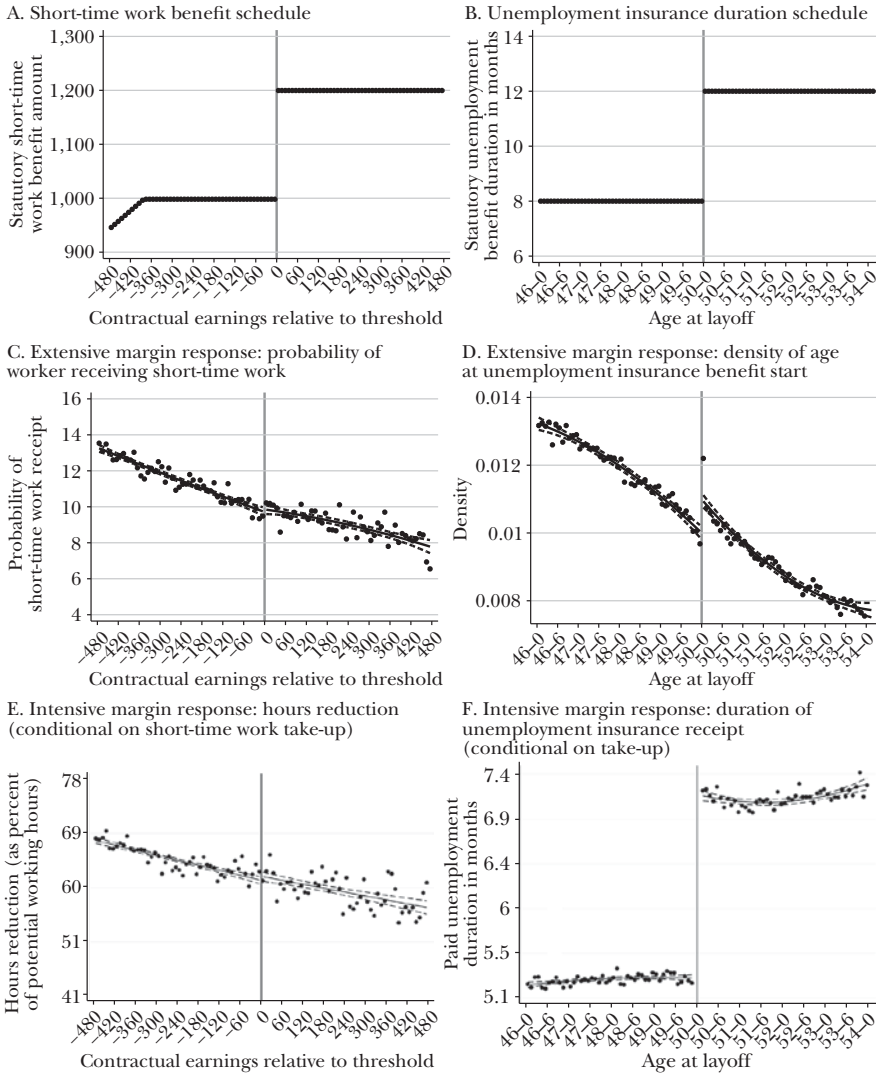
⁸Using data on firm balance sheets, Giupponi and Landais (2020b) show that in firms taking up short-time work, value-added per worker fell significantly, and by about the same magnitude as hours per worker. This indicates that reduction in hours upon take-up of short-time work is in large part a real response rather than a reporting response.

⁹In Italy, the amount of short-time work subsidy received by the worker is equivalent to 80 percent of forgone earnings due to hours not worked, up to a cap established by law each year. For example, in 2021, workers with contractual monthly earnings above €2,159.48 can receive benefits up to €998.19. For those with contractual earnings above €2,159.48, the cap is €1,119.72.

¹⁰The benefit amount is proportional to average wages earned in the three months before layoff up to a cap. Workers receive 60 percent of their average wage for the first six months, 50 percent for the following two months, and 40 percent for the remaining four months, if still eligible.

Figure 3

Short-Time Work versus Unemployment Insurance Generosity and Take-Up



Note: The figure reports a set of regression-discontinuity graphs to illustrate the relationship between short-time work/unemployment insurance generosity and take-up. Left-hand side panels report evidence on short-time work, right-hand side ones on unemployment insurance and are based on Scrutinio (2018). Panel A shows the short-time work benefit schedule as a function of a wage threshold. The benefit amount is based on a 70 percent hour reduction and 2021 short-time work cap parameters. Panel C reports the probability for a worker to be put on short-time work in a given year (pooling years 2011–2014), conditional on being employed in a firm that uses short-time work in that year. Panel E reports the average number of hours a worker spends on short-time work per month (as a percent of total potential working hours), conditional on being on short-time work in that month. Panel B plots unemployment insurance potential benefit duration as a function of a worker’s age at layoff. Panel D reports the density of layoffs by age at layoff. Panel F shows the average duration of unemployment insurance in months, conditional on unemployment insurance take-up. Panels C–F are based on Italian Social Security data.

Rigidities or frictions to individual level bargaining within the firm may explain why short-time work take-up does not respond much to variation in the generosity of the subsidy at the individual level. These rigidities in turn can also rationalize why firms' behavior is generally much more responsive to a variation in the firm's rather than the worker's side of the job surplus (Jäger, Schoefer, and Zweimüller 2019). Firms appear to be responsive to how social insurance programs like unemployment insurance or short-time work are financed: for instance, if unemployment insurance is funded by experience-rated premia on firms (that is, firms where more workers claim unemployment insurance pay somewhat higher premia), this tends to reduce layoffs significantly. Time-series evidence also suggests that the take-up of short-time work by firms declines significantly with the tightening of short-time work's experience rating.

The last reason why the fiscal cost of short-time work appears limited, and probably the main reason for it, is that the fiscal cost of short-time work is of course endogenous to the generosity of unemployment insurance. If more generous short-time work prevents layoffs, these positive employment effects mechanically reduce the fiscal cost to the unemployment insurance system, as fewer workers end up collecting unemployment insurance. In turn, the more generous the unemployment insurance system relative to short-time work, the larger will such savings be for the government. This leads us to the central question: does short-time work effectively save jobs? If so, what are the welfare consequences?

Short-Time Work and Job Destruction

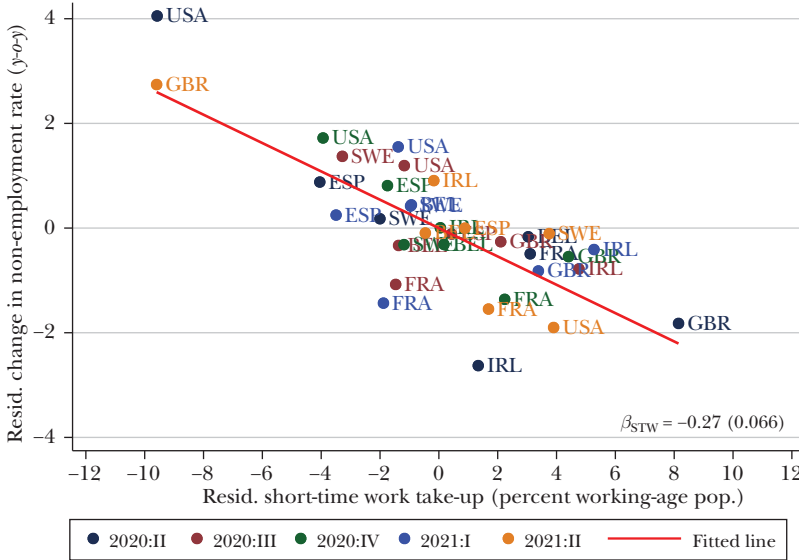
While both unemployment insurance and short-time work offer insurance against labor market shocks, they differ in a fundamental way. Short-time work seeks to preserve labor market matches by subsidizing jobs rather than job-seekers. In other words, short-time work, contrary to unemployment insurance, aims at reducing job separations. It is therefore critical to establish to what extent short-time work effectively saves jobs. Moreover, the social welfare impact of saving jobs will depend on whether separations are *inefficiently* high in recessions to begin with. So let us first review the evidence on the employment effects of short-time work, before delving into the reasons why layoffs may be inefficient and determining whether subsidizing labor hoarding may be socially desirable.

The Employment Effects of Short-Time Work

To determine whether short-time work saves jobs, a natural place to start is to leverage the large variation in short-time work usage during the COVID crisis across countries. Figure 4 provides such a macro perspective and shows the presence of a very robust negative correlation between the fraction of the working-age population that took up short-time work and the evolution of the non-employment rate during the crisis. One additional worker enrolling in short-time work is correlated with .27 fewer workers being non-employed. This strong correlation between employment and short-time work usage echoes time-series and cross-country evidence from previous recessions (for example, Van Audenrode

Figure 4

Short-Time Work Usage and Non-Employment During the COVID Crisis: Cross-Country Evidence



Note: The figure reports a scatter plot of the relationship between the year-on-year change in the quarterly non-employment rate and in the rate of short-time work take-up at the country level. Data are not seasonally adjusted. To remove the seasonal component, we take the year-on-year change—i.e., for a given year-quarter $YYQX$, we apply the following transformation to the data: $\tilde{x}_{YYQX} = x_{YYQX} - x_{(YY-1)QX}$. Short-time work take-up is computed as the ratio of the number of individuals in the program over the working-age population. For short-time work take-up in 2020 we take the variable in level as take-up was close to 0 in 2019. Outcomes are residualized against year-quarter fixed effects, the year-on-year change in the number of COVID cases (linear and quadratic), and in unemployment insurance take-up. The red line represents the linear fit. The figure reports the slope coefficient and associated standard error (in parentheses), clustered at the country level. Data on employment come from OECD. Data on short-time work and unemployment insurance take-up come from the OECD and national statistics (Scarpetta, Carcillo, and Hijzen 2022). Data on COVID cases come from the Johns Hopkins Coronavirus Resource Center. See online Appendix C for details on data sources and the construction of short-time work/unemployment insurance take-up.

1994; Abraham and Houseman 1993; Boeri et al. 2011; Hijzen and Venn 2011; Cahuc and Carcillo 2011).

Going beyond such correlations is complicated, and direct causal evidence on the employment effects of short-time work is scant.¹¹ The issue lies in the lack of credibly exogenous sources of variation in short-time work treatment across firms—an issue that will become even more acute for the current recession, as most countries have purposefully extended short-time work access to every firm. This situation severely complicates identification, with no obvious method to control for the selection of firms into short-time work.

¹¹ Using firm survey and administrative data in Denmark at the onset of the COVID pandemic, Bennedsen et al. (2020) compare actual furlough decisions and counterfactual decisions in the absence of aid. They find that the policy was effective in preserving job matches at the beginning of the pandemic.

Three recent papers focusing on the Great Recession do seek to address these selection problems and provide credible evidence of a positive, strong, and causal relationship between short-time work and employment. Looking at Swiss firms, Kopp and Siegenthaler (2021) compare firms whose short-time work application was granted to similar firms whose application was rejected. The unsuccessful establishments provide a valid counterfactual for the successful ones because approval practices across Swiss cantons are partly idiosyncratic. They find that short-time work prevented a large number of dismissals and significantly reduced the incidence of long-term unemployment.

Looking at French firms, Cahuc, Kramarz, and Nevoux (2021) use the proximity of a firm to other firms that used short-time work in the past as an instrumental variable for short-time work take-up during the Great Recession. As an alternative instrument, they use response-time variation in the administrative treatment of short-time work applications across French departments. They find large and significant employment effects of short-time work treatment.

Finally, looking at data from Italy, Giupponi and Landais (2020b) exploit plausibly exogenous variation in short-time work eligibility rules based on the interaction between industry and firm size. Their approach and main results are presented in Figure 5. The figure shows the evolution of the difference in short-time work take-up between eligible firms and similar counterfactual firms without access to short-time work, around the time of the Great Recession (blue diamonds). The evidence confirms that after the onset of the crisis, the take-up of short-time work among eligible firms surged quickly. The chart further reports the evolution of hours (red triangles) and total employment (purple circles) in eligible firms relative to counterfactual non-eligible firms. It demonstrates that short-time work had large and significant effects on firms' employment at both the intensive and extensive margin. Compared to counterfactual firms, firms treated by short-time work experienced a 40 percent reduction in hours worked per employee, which was met by an increase of similar magnitude in the number of headcount employees. Consistent with the findings of Kopp and Siegenthaler (2021) using Swiss firms, further results show that the employment effects are mostly driven by a reduction in dismissals among firms that would otherwise experience mass layoffs. Interestingly, Giupponi and Landais (2020b) also find no effect of short-time work on the wages of incumbents nor on the wages of new hires.

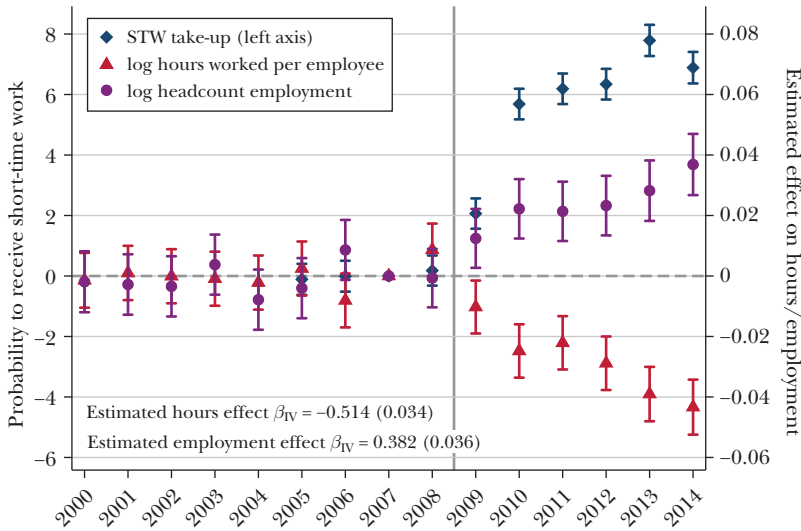
The Welfare Value of Labor Hoarding Subsidies

Overall, short-time work does seem to preserve jobs. But why is that valuable? In other words, why are employment adjustments at the intensive margin (hours reduction) versus extensive margin (layoffs) not equivalent in terms of welfare?

Preserving job matches is valuable for at least three reasons. First, frictions in the labor market, as well as hiring and training costs, make it costly for firms to replace workers and for workers to change jobs. Second, workers may accumulate human capital that is specific to their job, and separations risk destroying this valuable source of idiosyncratic productivity. Finally, unemployment often entails long-run scarring effects for workers (as discussed, for example, by Sullivan and

Figure 5

Employment Effects of Short-Time Work in Italy: Evidence from Quasi-random Eligibility Variation across Firms



Note: The figure reproduces estimates from Giupponi and Landais (2020b) on the effect of short-time work eligibility on the probability of short-time work take-up, the log of hours worked per employee, and the log of employment headcount at the firm level. The graph reports the estimated coefficients and associated confidence intervals (capped vertical bars) from a reduced-form regression of the outcome of interest on an indicator of short-time work-eligibility at the firm level interacted with year dummies. All results are relative to 2007. The graph also reports the instrumental variable coefficient (and standard errors in parentheses) of the effect of short-time work take-up on log hours worked per employee and log employment headcount.

von Wachter 2009). As a consequence, we should observe significant *labor hoarding*: firms and workers should be willing to preserve matches when hit by negative shocks.

But in practice, the socially efficient level of labor hoarding may not be achieved. Liquidity constraints are probably the most obvious and prevalent reason: during a shock, a firm may lack the funds necessary to pay wages and retain its workers. Giroud and Mueller (2017) document that, during the Great Recession, US firms facing higher liquidity constraints, as proxied by pre-crisis levels of leverage, were (all else equal) more likely to reduce employment in response to a consumer demand shock. Of course, employers could try to negotiate temporary wage or hours adjustments with their employees to deal with such liquidity constraints. But bargaining costs and commitment issues may often make such renegotiation impractical. Therefore, wage and hours rigidities may interact with liquidity constraints to amplify the employment response to negative shocks (Schoefer 2016; Jäger, Schoefer, and Zweimüller 2019). Finally, note that generous and imperfectly experience-rated unemployment insurance may also already distort workers' and firms' choices in favor of (socially inefficient) dismissals. As pointed out by Braun and Brügemann (2014), this interaction

between short-time work and the pre-existing distortions caused by unemployment insurance is critical to the welfare analysis of short-time work.

If separations are indeed inefficiently high during recessions because of liquidity constraints and other bargaining frictions, subsidizing labor hoarding can be efficient. Indeed, evidence from Giupponi and Landais (2020b) regarding Italian firms strongly supports this idea. It shows that liquidity-constrained firms, identified using various indicators from balance-sheet data, were much more likely to take up short-time work. Moreover, the treatment effects of short-time work were much more positive for these firms. The number of jobs saved per subsidized hour was significantly larger for them, and so was the effect of short-time work on the probability that the firm survived.

In sum, the liquidity constraint channel seems critical in explaining the excess sensitivity of employment adjustments to productivity shocks and supports the idea of having job match subsidies to correct for inefficiently high separations. Yet two important questions remain.

First, what is the welfare value of saving these jobs? The answer depends on the value of the surplus of the marginal job match saved: the larger the value of a match, the larger the positive welfare effect of preserving it. Unfortunately, this value is an object that is hard to fathom, let alone to measure precisely, and on which there is little consensus in the literature.

Second, why should short-time work be the only way to implement such subsidies? What about other policy instruments? A natural alternative instrument would be “recall” unemployment insurance, under which workers can return to their former employer after a spell on unemployment insurance, thus preserving the job match. But in practice, recall unemployment insurance entails much less commitment to preserving the job match. Furthermore, it lacks the flexibility to insure against partial reductions in hours, a flexibility which can prove effective in addressing financial constraints and in preserving employment. What about direct wage subsidies, or direct provision of liquidity with temporary loans, such as the Paycheck Protection Program in the United States? If liquidity constraints are in fact the main underlying source of inefficiency, tools addressing these financial constraints directly may be more appropriate than short-time work.

But even with these alternatives in mind, two arguments tend to support short-time work. The option of short-time work offers expediency: it can almost immediately provide the funds necessary to cover a firm’s payroll. In contrast, dedicated loan programs, as demonstrated by the experience of the US Paycheck Protection Program, can take more time to be activated and for funds to actually reach firms. Short-time work also may offer superior targeting, because it channels liquidity to firms that are willing to reduce their hours, which tends to be an effective screening mechanism. In practice, short-time work selects firms effectively hit by negative shocks, as measured by revenues, labor productivity, or the predicted probability to engage in mass layoffs (Giupponi and Landais 2020b; Kopp and Siegenthaler 2021). This screening property makes short-time work more effective than non-targeted wage subsidies, which can end up subsidizing a lot of non-marginal matches.

Reallocation in the Labor Market

Recessions are times of intense reallocation between workers and firms (Foster, Grim, and Haltiwanger 2016). They are also usually characterized by slackness in the labor market: many workers are searching for jobs while firms demand less labor and post fewer vacancies.

Both short-time work and unemployment insurance affect workers' search effort and firms' labor demand. To assess the social desirability of both programs, we must therefore also factor in how they impact labor market reallocation. We start by focusing on the effect of both social insurance schemes on the overall tightness of the labor market. We then delve into the impact of both policies on the optimal sorting of workers into firms in the labor market.

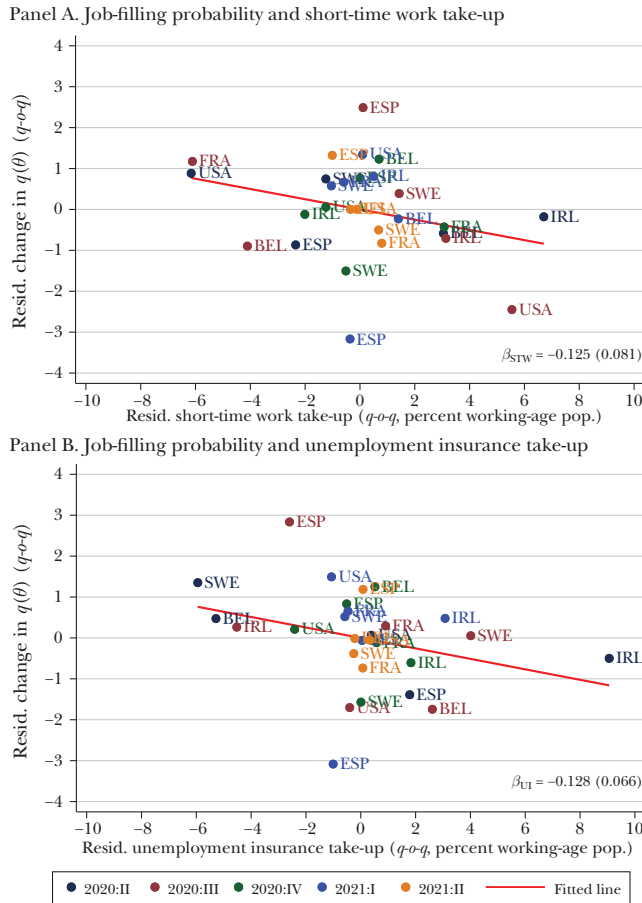
Social Insurance and the Tightness of the Labor Market

Reallocation in the labor market occurs as workers search for new jobs and firms hire new workers. When there are a lot of workers searching for jobs, it is easy for firms to hire new workers: the labor market is "slack," which is good for firms, but bad for workers. When, to the contrary, there are a lot of vacancies but few workers searching for jobs, the labor market is "tight."

Unemployment insurance and short-time work affect at the same time the aggregate search effort of workers and the incentives for firms to hire new workers: their overall effect on the tightness of the labor market is therefore a priori ambiguous. If generous unemployment insurance increases wages or if short-time work strongly reduces the need for new hires, more generous social insurance might strongly reduce the number of vacancies posted by firms and make the labor market even more slack in recessions, delaying recovery. To the contrary, if labor demand is rigid during recessions and the labor market exhibits job rationing, workers searching for jobs find themselves in a rat race. In such contexts, reducing search effort through more generous unemployment insurance or short-time work can benefit workers by increasing the tightness of the labor market (Landais, Michailat, and Saez 2018b).

What do the data tell us about the overall effect of unemployment insurance and short-time work on equilibrium tightness? As a starter, we can again exploit the large variation in short-time work and unemployment insurance usage across countries and over time during the recent crisis. For this purpose, we built consistent measures of job-filling probabilities, using the ratio of hires to vacancies. These measures are direct proxies of the slackness of the market: the tighter the market, the harder it is for firms to hire workers, and the lower the job-filling probability as a result. We then correlate the change in job-filling probabilities with the change in short-time work and in unemployment insurance take-up across countries and across quarters during the current recession. Our results, reported in Figure 6, show that short-time work (Panel A) and unemployment insurance (Panel B) both seem negatively correlated with job-filling probabilities (or equivalently positively correlated with labor market tightness) in a recession, which is consistent with the presence of significant job rationing in downturns. The evidence suggests that increasing the

Figure 6
Cross-Country Correlation between Job-Filling Probability and Take-Up of Short-Time Work and Unemployment Insurance



Note: The figure shows how short-time work and unemployment insurance take-up during the COVID crisis correlate with tightness in the labor market. We use the vacancy-filling probability $q(\theta)$ as a proxy for labor market tightness. The higher the vacancy-filling probability, the easier it is for firms to hire workers when opening a vacancy, and the slacker the labor market as a result. Both panels report scatter plots of the relationship between the quarter-on-quarter change in $q(\theta)$ and in the rate of short-time work/unemployment insurance take-up at the country level. Data are not seasonally adjusted. To remove the seasonal component, we take the quarter-on-quarter change—that is, for a given year-quarter $YYQX$, we apply the following transformation to the data: $\tilde{x}_{YYQX} = x_{YYQX} - x_{YYQ(X-1)}$. Short-time work and unemployment insurance take-up are computed as the ratio of the number of individuals in the program over the working-age population. Outcomes are residualized against year-quarter fixed effects, the quarter-on-quarter change in the number of COVID cases (linear and quadratic), and in the take-up of the other policy instrument. The red line represents the linear fit. The figure reports the slope coefficient and associated standard error (in parentheses), clustered at the country level. Data for European countries come from the Job Vacancy Statistics and Labor Force Survey, and for the United States from the Job Openings and Labor Turnover Survey. For European countries, hires are proxied by recent job starters—i.e., individuals who reported having started their employment in the last three months before the interview. Job openings are restricted to the private sector. Data on short-time work and unemployment insurance take-up come from the OECD and national statistics (Scarpetta, Carcillo, and Hijzen 2022). Data on COVID cases come from the Johns Hopkins Coronavirus Resource Center. See online Appendix C for details on data sources and the construction of short-time work/unemployment insurance take-up.

generosity of short-time work or unemployment insurance in a recession can be an effective way of alleviating the search inefficiencies created by rat-race externalities.

This cross-country evidence is corroborated by a stream of recent papers that identify the impact of social insurance on search externalities and equilibrium tightness using quasi-experimental designs. Lalive, Landais, and Zweimüller (2015) exploit a massive expansion in the generosity of unemployment insurance to a large subgroup of workers in Austria and show that non-eligible workers have significantly higher job finding rates, lower unemployment durations, and lower risk of long-term unemployment as a result. Marinescu (2017) uses job board data and exploits quasi-random variation in unemployment insurance expansions across US states during the Great Recession: she finds that unemployment insurance reduced search effort significantly but did not affect job vacancies, so that tightness went up significantly as a result. Marinescu, Skandalis, and Zhao (2020, 2021) exploit variation in unemployment insurance across US labor markets stemming from the CARES Act and Federal Pandemic Unemployment Compensation (FPUC). Using granular data from the online job platform Glassdoor, they show in both cases that increases in unemployment insurance generosity significantly increased labor market tightness. Finally, using exogenous variation across local labor markets in Italy in their exposure to short-time work, Giupponi and Landais (2020b) find that greater access to short-time work decreases the job finding probability in the labor market, but that the magnitude of the effect is small. Overall, these results confirm that both unemployment insurance and short-time work increase tightness during downturns, and the effect seems to be more pronounced for unemployment insurance.

Of course, the welfare consequences of increasing tightness depend on whether tightness is inefficiently low or high in recessions. Historically, labor markets tend to be very slack during downturns. Michaillat and Saez (2021) offer a general characterization as well as a measure of the efficient level of tightness in the United States. They find that the labor market has been particularly inefficiently slack during past recessions. The intuition is that the social cost of unemployment is very large relative to firms' recruiting costs during downturns. Pushing tightness up and increasing the job-finding probability of workers is then socially desirable: the reduction in the social cost of unemployment greatly outweighs the increased costs of recruiting for firms.

However, evidence from the current crisis suggests that this time *is* different. Looking at the long-run evolution of the average vacancy-filling probability in the United States, it is striking to see that it has remained at a historic low during the crisis. Overall, this recession seems unique: it is a tight recession in the labor market.¹²

Can this sustained level of tightness actually be explained by the large expansion of unemployment insurance generosity and coverage in the United States at the onset of the COVID crisis? Would the situation be different if US policymakers had resorted more to short-time work, which seems to put less upward pressure

¹²For more information on labor market tightness during the pandemic, including figures showing the hires-to-vacancies ratio for the US and various European economies, see online Appendix D. In the US data, the brief surge in the second quarter of 2020 can be entirely explained by early recalls from unemployment.

on tightness? When we consider comparable data for European countries, we find that they have also experienced what appears to be tight labor markets during the pandemic recession, which suggests that the mix of social insurance policies used during the COVID crisis is probably not responsible. But this pattern also implies that there is probably no need to push tightness further up going forward. Exploring the factors behind this uniquely high level of tightness during a slump is important to guide the policy response during the recovery.

Slowing Down Reallocation across Firms and Sectors

Recessions trigger shocks that are asymmetric across firms and sectors. As a result, significant reallocation usually follows in the labor market: workers move away from firms persistently hit by bad shocks and toward more productive job matches, a movement which enhances aggregate efficiency. In recent months, concerns have emerged again on the impact that higher social insurance might have on the pace of this sectoral and firm reallocation (for example, Barrero et al. 2021).

Both unemployment insurance and short-time work have the potential to hinder reallocation, although the mechanism by which they do so differs. In theory, unemployment insurance is a general brake to aggregate reallocation: by lowering the search effort of the unemployed, it can slow the pace at which workers who have been dismissed from lower productivity jobs may move to more productive matches. Short-time work is a specific brake to sectoral/firm reallocation: it discourages workers in firms/sectors that are hit by productivity shocks from reallocating to other firms/sectors by keeping them in their jobs. The extent to which this is problematic for aggregate productivity depends on whether the shock is temporary or permanent: if the shock is permanent, then short-time work may subsidize persistently unproductive matches and hinder reallocation towards more productive job matches.

How serious are these negative reallocation effects of unemployment insurance and short-time work in practice? Regarding unemployment insurance, we know surprisingly little on its overall impact on reallocation and aggregate efficiency in the labor market. For short-time work, evidence from Giupponi and Landais (2020b) on Italian firms sheds some interesting light on its impact on reallocation. It finds that short-time work tends to subsidize persistently low productivity matches, as low productivity firms tend to over-select into short-time work. Italian firms that were already below the median of labor productivity before the onset of the recession were twice as likely to select into short-time work during the Great Recession, and the employment effects of short-time work are also significantly lower for these low productivity firms. Furthermore, exploiting variation across local labor markets, they show that (exogenously) higher exposure to short-time work is significantly and negatively correlated with the employment growth of high productivity firms. In other words, high productivity firms have a harder time growing in a local labor market where low productivity firms have more access to short-time work. While this clearly supports the idea that short-time work slows down reallocation, the magnitude of the estimated effects remains small. However, the level of take-up of short-time work was also much smaller during the Great Recession than in the current crisis, and one

cannot exclude that short-time work may have much stronger negative effects on reallocation in the current recovery.

Further Externalities

Besides inefficient separations and reallocation frictions, it is worth pointing to a few further externalities with which short-time work and unemployment insurance may interact: aggregate demand externalities, fairness, and health.

A usual argument in favor of generous social insurance during recessions relates to their fiscal multiplier effects: unemployment insurance and short-time work transfer money to individuals who tend to have higher-than-average marginal propensities to consume. These high marginal propensities to consume, in turn, may help trigger positive aggregate demand externalities in a slump. A small literature has tried to embed social insurance into New Keynesian models to quantify the size of these multiplier effects (for example, McKay and Reis 2016; Michaillat and Saez 2019; Guerrieri et al. 2020; Kekre forthcoming). How large are these fiscal multiplier effects? Which program commands the larger fiscal multipliers: short-time work or unemployment insurance?

Unemployment insurance, as explained above, tends to insure individuals experiencing larger shocks with lower means to smooth consumption: this suggests that unemployment insurance recipients are likely to have larger marginal propensities to consume. But short-time work, by preserving employment and improving expectations regarding future employment and income, may reduce the need for precautionary savings and thus raise marginal propensities to consume compared to unemployment insurance.

The marginal propensities to consume of recipients of unemployment insurance are large, and significantly larger than those of employed people. Comparing the same individuals over time in Sweden, Landais and Spinnewijn (2021) find that the marginal propensity to consume is around 25 percent higher when unemployed than employed. But much less is known on the marginal propensities to consume of individuals on short-time work. We looked at the data from the German High-Frequency Online Personal Panel Survey (HOPP) mentioned earlier, and elicited marginal propensities to consume following the approach of Jappelli and Pistaferri (2014). We find that the marginal propensities to consume of German short-time work recipients was slightly larger than that of employed workers, but smaller than that of unemployment insurance recipients.¹³

However, moderate differences in marginal propensities to consume between unemployment insurance and short-time work recipients are unlikely to translate into sizeable differences in aggregate demand externalities between these two policies, because the fraction of the labor force receiving unemployment insurance or short-time work is small relative to the size of the employed population. For that

¹³For details and additional discussion, see online Appendix Table B1.

reason, simulations such as in McKay and Reis (2016) suggest that, quantitatively, the stabilization effects of these forms of social insurance are small and second-order.¹⁴

Fairness appears to be an important institutional tenet in European labor markets (Saez, Matsaganis, and Tsakoglou 2012; Goldschmidt and Schmieder 2017; Saez, Schoefer, and Seim 2019). Fairness concerns suggest that short-time work may prove a more equitable way to insure against labor market fluctuations. If firms avoid layoffs and instead reduce hours of work per worker, the costs of recessions are less concentrated on a small number of workers who suffer large losses in income and other job-related benefits. Interestingly, this argument is often mentioned in the policy debate in countries with strong short-time work programs.

Finally, in the current pandemic, the ability granted by short-time work to flexibly reduce hours of work and keep workers away from the workplace may have had positive health externalities by reducing the spread of the virus.

Conclusion

While very little was known about short-time work schemes and their potential welfare effects, this did not prevent European policymakers from resorting to them aggressively during the COVID crisis. The evidence gathered in this paper and summarized in Table 1 shows they probably did the right thing. The value of insurance provided by short-time work transfers is clearly lower than that of unemployment insurance benefits, but the moral hazard they entail seems more limited than for unemployment insurance, especially when experience-rating of social insurance is limited. Importantly, recent evidence confirms that short-time work can be an efficient and expedient way to attenuate the social costs created by “excess” layoffs in recessions.

But while the policy debate has tended to discuss a choice between short-time work and unemployment insurance, this paper has revealed two important points. First, the frontier between unemployment insurance and partial unemployment or short-time work policies can often be quite tenuous. In other words, in certain circumstances, unemployment insurance schemes can be made to mimic the functioning of short-time work or furlough schemes by maintaining important ties between the unemployed and their employers.

Second, far from being substitutes, these two types of policies exhibit strong complementarities. Unemployment insurance and short-time work do not insure the same types of workers nor the same labor market risks. They do not distort the same margins of behaviors. There are important fiscal spillovers between the two programs, as more generous short-time work reduces the risk of layoffs. Their effects on reallocation in the labor market are also complementary. Short-time work

¹⁴For social insurance to have large multiplier effects, it would need to have a strong effect on the consumption behavior of the large population of the employed as well. But in practice, the precautionary savings channel (by which employed individuals save less when they have access to more generous social insurance against labor market shocks) seems too small to sustain large aggregate demand externalities.

Table 1

The Welfare Effects of Insuring Workers versus Jobs: A Summary of the Evidence

	<i>Value of transfer</i>	<i>Moral hazard/fiscal externality</i>	<i>Correction of other inefficiencies</i>		
			<i>Excess layoff</i>	<i>Tightness extern.</i>	<i>Reallocation</i>
Short-time work	+	+/-	++	?	-
Unemployment insurance	++	--	?	+	?

Note: The table provides our own summary evaluation of the empirical literature evaluating features of short-time work and unemployment insurance programs that map onto those key elements of welfare analysis. For discussion of the underlying studies, see the text of the paper. The symbols reported in the table refer to the magnitude of the welfare effect for each feature, as per the following legend: (++) large positive, (+) positive, (+/-) both positive and negative, (-) negative, (--) large negative, (?) no evidence.

is an efficient policy to deal with temporary shocks, while unemployment insurance can take care of shocks that end up being more persistent. In countries with already generous unemployment insurance and/or strong employment protection, like European countries, strong cyclical short-time work programs are therefore a valuable complement to unemployment insurance as a policy response to recessions. And, in general, having both programs is a great way to handle any type of recession in the labor market. In other words, we should insure workers *and* jobs during recessions.

But this paper has also emphasized that social insurance critically interacts with equilibrium in the labor market, with important consequences for reallocation and efficiency. On this front, much more research needs to be done. As the current crisis seems to be unique in maintaining high tightness in the labor market, a better understanding of how unemployment insurance and short-time work affect reallocation will be key to determine the optimal policy path for the recovery. Attention should in particular be devoted to determining how unemployment insurance and short-time work should be coordinated with other instruments, such as hiring subsidies, in order to boost labor demand and prevent reallocation issues.

■ *We thank David Autor, Tito Boeri, Xavier Jaravel, Pascal Michaillat, Daniel Reck, Emmanuel Saez, and Johannes Spinnewijn, and seminar participants at CREST, the Institute for Economic Analysis at Universitat Autònoma de Barcelona, the SIEP Annual Conference, the International Monetary Fund and the Franco-German Seminar for helpful comments and suggestions. We are grateful to Andrea Salvatori and Alexander Hijzen for sharing OECD data on short-time work take-up, Robert Hall and Marianna Kudlyak for sharing data on recalls, and Vincenzo Scrutinio, Jan Tilly, and Kilian Nedermayer for sharing data and code from their papers. We also thank Francesco Armillei, Nicolas Grimprel, and Colombe Ladreit de Lacharrière for excellent research assistance. We gratefully acknowledge support from ERC grant #679704-DYNAMICSS, STICERD, ESRC, and the INPS “Valeria Solesin” Fellowship. The findings and conclusions expressed are solely those of the authors and do not represent the views of INPS.*

References

- Abraham, Katharine G., and Susan N. Houseman.** 1993. "Does Employment Protection Inhibit Labor Market Flexibility? Lessons from Germany, France, and Belgium." NBER Working Paper 4390.
- Albanese, Andrea, Matteo Picchio, and Corinna Ghirelli.** 2020. "Timed to Say Goodbye: Does Unemployment Benefit Eligibility Affect Worker Layoffs?" *Labour Economics* 65: 101846.
- Baily, Martin Neil.** 1978. "Some Aspects of Optimal Unemployment Insurance." *Journal of Public Economics* 10 (3): 379–402.
- Barrero, Jose Maria, Nicholas Bloom, Steven J. Davis, and Brent H. Meyer.** 2021. "COVID-19 Is a Persistent Reallocation Shock." University of Chicago, Becker Friedman Institute for Economics Working Paper 2021-02.
- Bennedsen, Morten, Birthe Larsen, Ian Schmutte, Daniela Scur.** 2020. "Preserving Job Matches During the COVID-19 Pandemic: Firm-Level Evidence on the Role of Government Aid." GLO Discussion Paper 588.
- Boeri, Tito, and Herbert Bruecker.** 2011. "Short-Time Work Benefits Revisited: Some Lessons from the Great Recession." *Economic Policy* 26 (68): 697–765.
- Braun, Helge, and Björn Brügemann.** 2014. "Welfare Effects of Short-Time Compensation." CESifo Working Paper Series 5063.
- Bundesagentur für Arbeit Statistik.** 2021. "Angezeigte und realisierte Kurzarbeit (Zeitreihe Monats und Jahreszahlen)." Bundesagentur für Arbeit, Germany. <https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Fachstatistiken/Leistungen-SGBIII/Kurzarbeitergeld/Kurzarbeitergeld-Nav.html> (accessed November 29, 2021).
- Bundesagentur für Arbeit Statistik.** 2021b. "Arbeitslosengeld - Deutschland, West/Ost, Länder, Kreise, Regionaldirektionen, Agenturen für Arbeit (Zeitreihe Monats- und Jahreszahlen ab 2005)." Bundesagentur für Arbeit, Germany. <https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Fachstatistiken/Leistungen-SGBIII/Arbeitslosengeld/Arbeitslosengeld-Nav.html> (accessed November, 29 2021).
- Bureau of Labor Statistics.** 2000–2021. "Job Openings and Labor Turnover Survey (JOLTS) - All estimates - jt.data.1.AllItems." Washington, DC: US Department of Labor. <https://download.bls.gov/pub/time.series/jt/jt.data.1.AllItems> (accessed November 27, 2021).
- Cahuc, Pierre, and Stéphane Carcillo.** 2011. "Is Short-Time Work a Good Method to Keep Unemployment Down?" *Nordic Economic Policy Review* 1 (1): 133–65.
- Cahuc, Pierre, Francis Kramarz, and Sandra Nevoix.** 2021. "The Heterogeneous Impact of Short-Time Work: From Saved Jobs to Windfall Effects." CEPR Discussion Paper 16168.
- Chetty, Raj.** 2008. "Moral Hazard Versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy* 116 (2): 173–234.
- Dares.** 2020. "Données trimestrielles – Activité partielle au 1er trimestre 2019." Ministère du Travail, France. <https://dares.travail-emploi.gouv.fr/donnees/le-chomage-partiel> (accessed November 29, 2021).
- Dares.** 2021a. "Données mensuelles – Données administratives de l'activité partielle (DAP, DI, heures, montants...)." Ministère du Travail, France. <https://dares.travail-emploi.gouv.fr/donnees/le-chomage-partiel> (accessed November 29, 2021).
- Dares.** 2021b. "Données mensuelles CVS-CJO - Demandeurs d'emploi." Ministère du Travail, France. <https://dares.travail-emploi.gouv.fr/donnees/les-demandeurs-demploi-inscrits-pole-emploi-france-metro> (accessed November 30, 2021).
- Dong, Ensheng, Hongru Du, Lauren Gardner.** 2020. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time." *Lancet* 20 (5): 533–4.
- Dun & Bradstreet.** 2021a. "ISO 2-Digit Alpha Country Code." Jacksonville: Dun & Bradstreet. https://www.dnb.com/content/dam/english/dnb-solutions/sales-and-marketing/iso_2digit_alpha_country_codes.xls (accessed April 7, 2021).
- Dun & Bradstreet.** 2021b. "ISO 3-Digit Alpha Country Code." Jacksonville: Dun & Bradstreet. https://www.dnb.com/content/dam/english/dnb-solutions/sales-and-marketing/iso_3digit_alpha_country_codes.xls (accessed April 7, 2021).
- Employment and Training Administration.** 2021. "UI Weekly Continued Claims – All programs." Washington, DC: Department of Labor. <https://oui.doleta.gov/unemploy/docs/allprograms.xls> (accessed November 29, 2021).
- Eurostat.** 2001–2021. "Job vacancy statistics by NACE Rev. 2 activity - quarterly data (from 2001 onwards)."

- Luxembourg: European Union Statistical Office. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=jvs_q_nace2&lang=en (accessed November 27, 2021).
- Eurostat.** 2001–2021b. “Job vacancy statistics by NACE Rev. 2 activity - quarterly data (from 2001 onwards).” Luxembourg: European Union Statistical Office. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=jvs_q_nace2&lang=en (accessed November 27, 2021).
- Eurostat.** 2009–2021. “Recent job starters by sex and age - quarterly data, AGE From 20 to 64 years, S_ADJ Unadjusted data (i.e. neither seasonally adjusted nor calendar adjusted data), SEX Total, UNIT Total - lfsi_sta_q.” Luxembourg: European Union Statistical Office. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsi_sta_q&lang=en (accessed November 27, 2021).
- Foster, Lucia, Cheryl Grim, and John Haltiwanger.** 2016. “Reallocation in the Great Recession: Cleansing or Not?” *Journal of Labor Economics* 34 (S1): 293–331.
- Giroud, Xavier, and Holger M. Mueller.** 2017. “Firm Leverage, Consumer Demand, and Employment Losses During the Great Recession.” *Quarterly Journal of Economics* 132 (1): 271–316.
- Giupponi, Giulia, and Camille Landais.** 2020a. “Building Effective Short-Time Work Schemes for the COVID-19 Crisis.” *Vox EU*, April 1.
- Giupponi, Giulia, and Camille Landais.** 2020b. “Subsidizing Labor Hoarding in Recessions: The Employment and Welfare Effects of Short Time Work.” CEPR Discussion Paper 13310.
- Giupponi, Giulia, and Camille Landais.** 2020c. “Replication Files for Figures 4 and 9.C: Subsidizing Labor Hoarding in Recessions: The Employment and Welfare Effects of Short Time Work.” CEPR Discussion Paper 13310.
- Goldschmidt, Deborah, and Johannes F. Schmieder.** 2017. “The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure.” *Quarterly Journal of Economics* 132 (3): 1165–217.
- Gruber, Jonathan.** 1997. “The Consumption Smoothing Benefits of Unemployment Insurance.” *American Economic Review* 87 (1): 192–205.
- Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning.** 2020. “Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?” NBER Working Paper 26918.
- Haas, Georg-Christoph, Bettina Müller, Christopher Osiander, Julia Schmidtke, Annette Trahms, Marieke Volkert, and Stefan Zins.** 2021. “Development of a New COVID-19 Panel Survey: The IAB High-Frequency Online Personal Panel (HOPP).” *Journal for Labour Market Research* 55 (16).
- Hendren, Nathaniel.** 2017. “Knowledge of Future Job Loss and Implications for Unemployment Insurance.” *American Economic Review* 107 (7): 1778–823.
- Hijzen, Alexander, and Danielle Venn.** 2011. “The Role of Short-Time Work Schemes During the 2008-09 Recession.” OECD Social, Employment Migration Working Paper 115.
- INPS.** 2021. *Report Mensile Novembre 2021 – Cassa integrazione guadagni e Disoccupazione*. Istituto Nazionale di Previdenza Sociale, Italy. <https://www.inps.it/osservatoristatistici/5> (accessed on December 1, 2021).
- INPS.** 2021b. “Visitinps Scholar Program Data.” Istituto Nazionale di Previdenza Sociale, Italy. <https://www.inps.it/dati-ricerche-e-bilanci/attivita-di-ricerca/programma-visitinps-scholars>.
- Jäger, Simon, Benjamin Schoefer, and Josef Zweimüller.** 2019. “Marginal Jobs and Job Surplus: A Test of the Efficiency of Separations.” NBER Working Paper 25492.
- Jappelli, Tullio, and Luigi Pistaferri.** 2014. “Fiscal Policy and MPC Heterogeneity.” *American Economic Journal: Macroeconomics* 6 (4): 107–36.
- Kekre, Rohan.** Forthcoming. “Unemployment Insurance in Macroeconomic Stabilization.” *Review of Economic Studies*.
- Khoury, Laura.** 2021. “Unemployment Benefits and the Timing of Redundancies.” Unpublished.
- Kopp, Daniel, and Michael Siegenthaler.** 2021. “Short-Time Work and Unemployment in and After the Great Recession.” *Journal of the European Economic Association* 19 (4): 2283–321.
- Lalive, Rafael, Camille Landais, and Joseph Zweimüller.** 2015. “Market Externalities of Large Unemployment Insurance Extension Programs.” *American Economic Review* 105 (12): 3564–96.
- Landais, Camille, Pascal Michailat, and Emmanuel Saez.** 2018a. “A Macroeconomic Approach to Optimal Unemployment Insurance: Applications.” *American Economic Journal: Economic Policy* 10 (2): 182–216.
- Landais, Camille, Pascal Michailat, and Emmanuel Saez.** 2018b. “A Macroeconomic Approach to Optimal Unemployment Insurance: Theory.” *American Economic Journal: Economic Policy* 10 (2): 152–81.
- Landais, Camille, and Johannes Spinnewijn.** 2021. “The Value of Unemployment Insurance.” *Review of Economic Studies* 88 (6): 3041–85.

- Marinescu, Ioana Elena.** 2017. “The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board.” *Journal of Public Economics* 150: 14–29.
- Marinescu, Ioana Elena, Daphné Skandalis, and Daniel Zhao.** 2020. “Job Search, Job Posting and Unemployment Insurance During the COVID-19 Crisis.” Unpublished.
- Marinescu, Ioana Elena, Daphné Skandalis, and Daniel Zhao.** 2021. “The Impact of the Federal Pandemic Unemployment Compensation on Job Search and Vacancy Creation.” *Journal of Public Economics* 200: 104471.
- McKay, Alisdair, and Ricardo Reis.** 2016. “The Role of Automatic Stabilizers in the U.S. Business Cycle.” *Econometrica* 84 (1): 141–94.
- Michaillat, Pascal, and Emmanuel Saez.** 2019. “Optimal Public Expenditure with Inefficient Unemployment.” *Review of Economic Studies* 86 (3): 1301–31.
- Michaillat, Pascal, and Emmanuel Saez.** 2021. “Beveridgean Unemployment Gap.” *Journal of Public Economics Plus* 2: 100009.
- Nomis Official Labour Market Statistics.** 2021. “Claimant count by sex and age.” Office for National Statistics, United Kingdom. <https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=162> (accessed on November 30, 2021).
- OECD.** 2021. “Social Benefit Recipients - High Frequency Database (SOCR-HF)” OECD, Paris. <https://www.oecd.org/els/soc/recipients-socr-hf.htm> (accessed on November 30, 2021).
- OECD.** 2021b. “Labour Force Statistics – Short-Term Labour market Statistics.” OECD, Paris. <https://stats.oecd.org/> (accessed on December 29, 2021).
- OECD.** 2022. *Riding the Waves: Adjusting Job Retention Schemes through the COVID-19 Crisis*. Paris: OECD Publishing.
- Saez, Emmanuel, Manos Matsaganis, and Panos Tsakloglou.** 2012. “Earnings Determination and Taxes: Evidence from a Cohort-Based Payroll Tax Reform in Greece.” *Quarterly Journal of Economics* 127 (1): 493–533.
- Saez, Emmanuel, Benjamin Schoefer, and David Seim.** 2019. “Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers’ Tax Cut in Sweden.” *American Economic Review* 109 (5): 1717–63.
- Scarpetta, Stefano, Stéphane Carcillo, and Alexander Hijzen.** 2022. *Riding the Waves: Adjusting Job Retention Schemes through the COVID-19 Crisis*. Paris: OECD Publishing.
- Scarpetta, Stefano, Mark Pearson, Alexander Hijzen, and Andrea Salvatori.** 2020. “Job Retention Schemes During the COVID-19 Lockdown and Beyond.” OECD Policy Responses to Coronavirus (COVID-19). October 12. https://read.oecd-ilibrary.org/view/?ref=135_135415-6bardplc5q&title=Job-retention-schemes-during-the-COVID-19-lockdown-and-beyond.
- Schmieder, Johannes F., Till von Wachter, and Stefan Bender.** 2016. “The Effect of Unemployment Benefits and Nonemployment Durations on Wages.” *American Economic Review* 106 (3): 739–77.
- Schoefer, Benjamin.** 2016. “The Financial Channel of Wage Rigidity.” Society for Economic Dynamics 2016 Meeting Papers 1605.
- Scrutinio, Vincenzo.** 2018. “The Medium-Term Effects of Unemployment Benefits.” Istituto Nazionale Previdenza Sociale (INPS) Working Paper 18.
- Scrutinio, Vincenzo.** 2018b. “Replication Files for Figures 1 and 3: The Medium-Term Effects of Unemployment Benefits.” INPS Working Paper 18.
- Sullivan, Daniel G., and Till von Wachter.** 2009. “Job Displacement and Mortality: An Analysis Using Administrative Data.” *Quarterly Journal of Economics* 124 (3): 1265–306.
- Tilly, Jan, and Kilian Niedermayer.** 2016. “Employment and Welfare Effects of Short-Time Work.” Unpublished.
- Tilly, Jan, and Kilian Niedermayer.** 2016b. “Replication Files for Figure 2.A: Employment and Welfare Effects of Short-Time Work.” Available upon request (contact: jantilly@gmail.com).
- Van Audenrode, Marc A.** 1994. “Short-Time Compensation, Job Security, and Employment Contracts: Evidence from Selected OECD Countries.” *Journal of Political Economy* 102 (1): 76–102.
- Volkert, Marieke, Georg-Christoph Haas, Stefan Zins, Lisa Hellmann, Sandra Dummer, Sophie Hensgen, Johannes Ludsteck, et al.** 2021. “High-frequency Online Personal Panel (IAB-HOPP).” Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). DOI: 10.5164/IAB.HOPP_W01-W07.de.en.v2

The \$800 Billion Paycheck Protection Program: Where Did the Money Go and Why Did It Go There?

David Autor, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz

In the early weeks of the COVID-19 pandemic, many small businesses in the United States were in precarious financial condition: revenues had plunged, access to credit was in many cases inadequate or absent, and large-scale layoffs and closures had already occurred (Bartik et al. 2020a,b). The potential consequences of widespread business failure were not confined to business owners. Since approximately 47 percent of US workers were employed by small businesses prior to the pandemic (SBA 2019), these closures held the potential for vast job loss. Over the longer term, widespread firm closures could slow the subsequent economic recovery by destroying intangible firm capital, liquidating high quality worker-firm matches, and forcing the costly reallocation of physical capital.

To aid these distressed businesses, Congress enacted the Paycheck Protection Program (PPP), which provided uncollateralized, low-interest loans of up to \$10 million to firms with fewer than 500 employees—loans that were forgivable on the condition that recipient firms maintained employment and wages at close

■ *David Autor is Professor of Economics, Massachusetts Institute of Technology, and also a Research Associate, National Bureau of Economic Research, both in Cambridge, Massachusetts. David Cho is an Economist, Leland Crane is a Senior Economist, Byron Lutz is an Assistant Director, Joshua Montes is a Senior Economist, William Peterman is Chief of the Fiscal Analysis Section, David Ratner is a Principal Economist, and Daniel Villar is an Economist, all at the Board of Governors of the Federal Reserve System, Washington, DC. Mita Goldar is an independent researcher. Ahu Yildirmaz is CEO of the Coleridge Initiative, Chevy Chase, Maryland. David Autor is the corresponding author at dautor@mit.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.55>.

to pre-crisis levels in the two to six months following loan receipt. The scale of the aid provided was extraordinary. By the time the program concluded in mid-2021, around \$800 billion in loans had been extended. Despite facing initial capacity constraints, the Paycheck Protection Program was notably successful in distributing a vast number of loans in short order: the take-up rate among eligible firms was 94 percent. Crucial to this rapid rollout was the decision to enlist the private sector to oversee the origination of all PPP loans, with the Small Business Administration (SBA) serving as the guarantor.

The Paycheck Protection Program was ultimately comparable in size to the two other major federal transfer programs enacted in response to the pandemic: expenditures on household payments—that is, stimulus checks—were around \$800 billion; and expenditures on expanded unemployment benefits totaled roughly \$680 billion under the Federal Pandemic Unemployment Compensation program (FPUC), Pandemic Unemployment Assistance program (PUA), and Pandemic Emergency Unemployment Compensation (PEUC) (CRFB 2021). As another standard of comparison, each of these three programs was roughly comparable in size to the *entire* American Recovery and Reinvestment Act of 2009 (ARRA), the principal fiscal stimulus enacted in response to the Great Recession of 2007–2009.

This paper explores who ultimately benefited from those \$800 billion in Paycheck Protection Program loans: concretely, where did the money go and why did it go there? We provide an answer in three steps. First, we consider how PPP funds flowed to three proximate sets of actors: workers who otherwise would have been laid off; creditors and suppliers of PPP-receiving businesses (for example, landlords, and utilities) who would otherwise not have received payments; and windfall transfers to PPP-recipient businesses (owners and shareholders) that would have maintained employment and met other financial obligations absent the PPP. Second, we calculate how these recipients were distributed across the household income distribution. Finally, we compare this allocation of funds to the household incidence of the two other major federal pandemic transfer programs: unemployment assistance and direct household payments. Our analysis combines lessons from existing research, including some of our own, and also presents new analysis using anonymized and aggregated payroll data from the private firm ADP, which processes payrolls for over 26 million individual workers in the United States per month.

The Paycheck Protection Program had measurable impacts. It meaningfully blunted pandemic job losses, preserving somewhere between 1.98 and 3.0 million job-years of employment during and after the pandemic at a substantial cost of \$169,000 to \$258,000 per job-year saved. PPP also reduced the rate of temporary closures among small firms, though it is less clear whether it reduced permanent closures. The majority of PPP loan dollars issued in 2020—66 to 77 percent—did *not* go to paychecks, however, but instead accrued to business owners and shareholders. And because business ownership and share-holding are concentrated among high-income households, the incidence of the program across the household income distribution was highly regressive. We estimate that about

three-quarters of PPP benefits accrued to the top quintile of household income. By comparison, the incidence of federal pandemic unemployment insurance and household stimulus payments was far more equally distributed.

Ironically, the program feature that arguably made the Paycheck Protection Program's meteoric scale-up possible is also the feature that made it potentially the most problematic: the program was essentially untargeted, aside from excluding firms with more than 500 workers (a rule further relaxed for some sectors). Small firms merely needed to attest that they were "substantially affected by COVID-19" to qualify, and almost all did so. Evidence strongly suggests that the program did not ultimately differentiate among firms or geographic areas according to need. This near absence of targeting virtually guaranteed that a large fraction of the first two tranches of \$525 billion in PPP loan dollars went to businesses that would have remained viable and retained their employees even absent PPP. Perhaps recognizing this program limitation, Congress explicitly targeted the final tranche (\$285 billion) of PPP loans in 2021 toward firms that had experienced revenue losses.

The Paycheck Protection Program's meteoric scale-up, its lack of targeting, and its highly regressive incidence reflect a key tradeoff that policymakers faced in March 2020 when crafting an emergency pandemic business loan program under severe time constraints: a lack of existing administrative infrastructure for overseeing large-scale, targeted federal support to US small businesses. Congress accordingly authorized the Small Business Administration (SBA) to harness the private sector to originate forgivable PPP loans and stipulated only a few coarse limitations on which firms could receive loans. These decisions rapidly opened the PPP floodgates to essentially all firms with fewer than 500 employees. Had policymakers instead insisted on better targeting, this would have likely substantially slowed aid delivery and reduced program efficacy. A key takeaway from the PPP experience is that building administrative capacity now would enable greatly improved targeting of either employment or business liquidity when the next pandemic or other large-scale economic emergency occurs, as it surely will.

The Basics

The Paycheck Protection Program sought to issue forgivable loans to small firms facing financial distress.¹ Businesses were permitted to draw PPP loans worth up to ten weeks of payroll costs—including wage and salary compensation not to exceed \$100,000 per worker, as well as paid leave, health insurance costs, other benefit costs, and state and local taxes—with a maximum loan size of \$10 million dollars.

¹The Paycheck Protection Program was one of four large government direct-lending programs introduced during the pandemic; the other three programs were the Main Street Lending Program, Corporate Credit Facilities, and Municipal Liquidity Facility. These programs jointly covered a large swath of the US economy (Decker et al. 2021).

Although the Small Business Administration issued the loan guarantees and would ultimately determine whether loans would be forgiven, PPP loans were processed and delivered through the nation's banking system.

The program received three tranches of funding. The Coronavirus Aid, Relief, and Economic Security Act of 2020 (CARES) established the Paycheck Protection Program and provided \$350 billion in appropriations on March 27, 2020. Subsequently, the Paycheck Protection Program and Health Care Enhancement Act, which passed on April 24, 2020, provided an additional \$320 billion in appropriations. A third tranche of \$285 billion was signed in to law on December 27, 2020, as part of the Consolidated Appropriations Act of 2021. Finally, early on in the pandemic, the Federal Reserve introduced the Paycheck Protection Program Liquidity Facility (PPLF) to bolster the ability of the banking system to provide PPP loans (Anbil et al. 2021).

Loans from the first two tranches were issued in 2020 and available to firms meeting the definition of a small business in the Paycheck Protection Program. In most industries, but not all, this required having fewer than 500 employees. The third tranche provided loans to firms in 2021 that had not previously taken out a PPP loan. It also provided "second draw" loans for firms that had already taken out a PPP loan, had fewer than 300 employees, and had experienced a significant revenue loss in 2020. About 75 percent of the third tranche of funding went to second-draw loans.

While the moniker Paycheck Protection Program suggests that the program was focused solely on employment, the criteria for loan forgiveness reveal another complementary goal: providing firms with liquidity to meet non-compensation obligations to creditors (like suppliers, banks, and landlords). Businesses had to do four things to qualify for PPP loan forgiveness: 1) spend at least 60 percent of the loan amount on payroll expenses; 2) spend (at least) the full loan amount on total qualifying expenses, including payroll, utilities, rent, and mortgage payments; 3) maintain average full-time equivalent employment at its pre-crisis level; and 4) maintain employee wages at no lower than 75 percent of their pre-crisis level. These criteria applied to a "covered period" that started on the date of loan disbursement and ran for 8 to 24 weeks, with the interval at the firm's discretion.

If these criteria were not met, the Small Business Administration offered alternative routes to forgiveness. Businesses could exercise a "safe harbor" option to meet the employment and wage criteria by restoring their full-time equivalent employment and wage rates to their pre-COVID level by the end of 2020 (or by the end of the covered period for loans issued in 2021). A second safe harbor absolved firms of the need to restore full pre-pandemic employment levels if they could document in "good faith" that other pandemic provisions (for example, lockdowns) made it infeasible to return to full business activity. These safe harbor provisions made the employment criteria far less onerous. Moreover, firms that did not meet all criteria could also receive partial loan forgiveness. As of late 2021, 94 percent of PPP loans issued in 2020 had applied for forgiveness and virtually all such applications had been approved by the Small Business Administration (SBA 2021).

One reason that almost all firms were able to meet these criteria is that they were retroactively loosened in June 2020, well after most PPP loans were issued (the discussion above pertains to the revised rules). Adding to the windfall, Congress amended the tax treatment of PPP loans in January 2021 to enable businesses to claim deductions for expenses paid with PPP loans (for example, wages, rent, utilities, etc.) without treating PPP loans as taxable business revenue. This retroactive change, which cost the Treasury an estimated \$100 billion in foregone tax revenue, effectively allowed some firms to pay a negative tax rate on PPP income (Harney and Mott 2021). For simplicity, our primary distributional accounting exercise below does not adjust for the additional tax subsidy provided to PPP recipients through this provision, though we briefly estimate its distributional implications—which are highly regressive.

A Timeliness versus Targeting Tradeoff

Fiscal interventions during economic downturns are often judged based on whether they are targeted, timely, and temporary (Elmendorf and Furman 2008). The Paycheck Protection Program was clearly *temporary*. How did it do on the other two T's?

Timeliness

The program deserves high marks for timeliness. When the pandemic began, no existing federal program had the scale to quickly distribute hundreds of billions of dollars to small businesses. The only other possible mechanism seemed to be state unemployment insurance systems (Bernstein and Rothstein 2020), but these systems struggled to handle the flood of initial unemployment insurance claims, and struggled further when tasked with distributing the enhanced unemployment benefits provided by the Coronavirus Aid, Relief, and Economic Security Act of 2020. It seems unlikely that state unemployment insurance systems could have handled an additional novel burden (Hubbard and Strain 2020).

Despite these obstacles, the Paycheck Protection Program succeeded in delivering a staggering sum of money over a two-month period in spring 2020. This can be seen in Table 1. As shown in column 3 of panel A, \$505 billion in first draw loans were issued to firms with fewer than 500 employees (column 3) and all but 7 percent of these were issued in 2020 (column 6). A very large share of these loans were issued in April and May (not shown). Finally, the memo lines show that non-employer businesses—for example, the self-employed—received \$43 billion in first draw loans and employers with more than 500 employees received a relatively small \$18 billion.

One emblem of the Paycheck Protection Program's success is its market penetration, which we define as the employment-weighted share of firms that received PPP loans and will refer to as the takeup rate. We make use of loan-level data from the PPP on the size of each firm that received a PPP loan, along with publicly available employment data from the Census Bureau's Statistics of US Businesses

Table 1
PPP Loans by Employer Size

<i>Employer size</i> (1)	<i>Employment share</i> (2)	<i>Loan \$ (billions)</i> (3)	<i>Share of \$</i> (4)	<i>Take-up rate</i> (5)	<i>% of \$ received in 2020</i> (6)
<i>A. First draw loans</i>					
1–4	10%	44	9%	81%	84%
5–9	11%	54	11%	98%	93%
10–49	35%	182	36%	99%	96%
50–149	23%	122	24%	97%	98%
150–299	13%	64	13%	91%	98%
300–499	9%	40	8%	87%	97%
1–499	100%	505	100%	94%	93%
<i>Memo:</i>					
Non-employers	–	43	–	–	25%
Employers 500+	–	18	–	–	93%
<i>B. Second draw loans</i>					
1–4	10%	17	9%	30%	–
5–9	11%	25	13%	43%	–
10–49	35%	87	46%	45%	–
50–149	23%	45	24%	34%	–
150–299	13%	14	8%	29%	–
300–499	9%	1	0%	3%	–
1–499	100%	189	100%	34%	–
<i>Memo:</i>					
Non-employers	–	11	–	–	–
Employers 500+	–	0.2	–	–	–

Source: Authors' analysis of Census Bureau Statistics of US Businesses (SUSB) 2018, BLS BED, and SBA PPP data.

Note: Panels A and B reflect data on employer businesses. The main panels exclude loans to the self-employed, sole proprietors, independent contractors, and single-member LLCs with only one reported job because non-employers are excluded from the SUSB data used to calculate the denominator of the takeup rates displayed in column 5. The roughly 4.6 million non-employer loans (constituting about 8 percent of total loan dollars) are reported in the first memo lines of each panel. As PPP loan-level data censor firm size at 500, in the main panels of the table we restrict attention to loans to businesses smaller than 500; loans to businesses reported as having 500 employees in the PPP loan-level data are reported in the second memo line of each panel. Loans to businesses in Guam, Puerto Rico, and the Virgin Islands are excluded. Loans to businesses in the following NAICS industries are excluded as they are out of scope for the SUSB data used in columns 2 and 5: 111, 112, 482, 491, 525110, 525120, 525190, 525920, 541120, 814, and 92.

(SUSB). SUSB data provide total employment for a number of categories of firm size which we use to form the rows of Table 1. For each size category, the takeup rate is the ratio of the total number of employees at PPP-receiving firms from the PPP loan data divided by total employment from SUSB. For example, in the PPP loan data, in the size bin 10–49, there were 1.3 million first-draw loans to firms with a total of 21.4 million employees over 2020 and 2021. In the aggregate, the SUSB data from 2018 (the latest available) report that there were 21.4 million employees in firms with between 10 and 49 workers; accounting for the growth

of employment between 2018 and before the pandemic, aggregate employment between 10 and 49 was 21.7 million. Thus, the takeup rate in this group is 99 percent (21.4 million/21.7 million), as given in column 5. We note that these estimated takeup rates are constrained by significant data limitations in determining the set of firms eligible for a PPP loan, inaccuracies in the reporting of firm size in PPP loan-receipt data, the possibility of fraudulent loans, and other measurement issues. (See the online Appendix, available with this article at the *JEP* website, for further details on the methodology underlying Table 1, as well as additional information on the subsequent analysis in this paper.)

Overall, we estimate that 94 percent of employers with fewer than 500 employees took up a Paycheck Protection Program loan; consistent with this high takeup rate, the distribution of loan dollars is tightly in line with employment shares—compare columns 2 and 4. Indeed, the fact that the second tranche of PPP funding concluded without exhausting all available funds suggests that the program had achieved something close to saturation in its first five months of operation. While near-universal participation in a government program is not altogether surprising since the program in most cases constituted a pure cash transfer, it is nevertheless a substantial administrative accomplishment: merely handing out \$500 billion dollars in two months takes many hands. As noted above, this accomplishment would likely have been infeasible had Congress not authorized the Small Business Administration to enlist the private banking sector to issue PPP loans.

The early rollout of the Paycheck Protection Program in April and May 2020 did, however, stumble on two hurdles. First, initial demand for loans significantly exceeded the ability of banks to deliver them. In the face of these capacity constraints, banks appear to have prioritized firms with which they had a preexisting relationship (Amiram and Rabetti 2020; Cororaton and Rosen 2021; Joaquim and Netto 2021; Granja et al. 2020; Li and Strahan, 2020). Larger firms, which tend to have ongoing banking relationships, accessed PPP funds sooner than smaller firms on average. Moreover, as most small business lending is sourced from local banks (Brevoort et al. 2010), the aptitude and willingness of local banks to process loan applications generated significant geographic heterogeneity in the initial distribution of loans (Bartik et al. 2021; Li and Strahan 2020).

The second hurdle was the significant uncertainty and confusion among businesses and banks over the specifics of the program, particularly over whether the loans would be forgiven. For example, in April 2020, the Small Business Administration announced that publicly traded companies were unlikely to satisfy the required good faith certification of need for a loan from the Paycheck Protection Program and stipulated a time window in which firms could return loans. Simultaneously, the Treasury Department announced that loans in excess of \$2 million would be subject to review and warned of possible criminal charges for those who failed the review. These issues were resolved over the course of several months. By the second round of funding, confusion about eligibility and forgiveness terms had abated. Meanwhile, initially underperforming banks upped their loan tempo, and non-banks stepped into fill gaps in local loan provision (Granja et al. 2020; Erel

and Liebersohn 2020). By July 2020, virtually all firms that would access a PPP loan in 2020 had done so.

The delay in delivering funds in April and May 2020 had real consequences. Doniger and Kay (2021) and Kurmann et al. (2021) find that loans received even a little earlier had a more pronounced effect on employment than those issued a bit later. Meanwhile, as we show below, the third tranche of loans, which did not go out until 2021, had no discernible effect on employment, perhaps because this tranche was issued when the labor market was already rapidly recovering.

Targeting

The rapid, near-universal takeup of Paycheck Protection Program loans in 2020 is inseparable from the reality that the program was essentially untargeted. That takeup was around 94 percent of *all* small businesses means that loans reached the most and least distressed firms—and all those in between—in nearly equal proportions. This observation helps to explain why there is little geographic correlation between the size of the initial COVID local economic shock, prior to PPP's passage, and subsequent PPP participation (Granja et al. 2020).

Around \$200 billion in so-called second draw loans were issued in 2021—see column 3 of Table 1, panel B. Unlike the first two tranches of Paycheck Protection Program funds, these loans were explicitly targeted at firms that had experienced significant revenue losses over the course of the pandemic (and had already received a first PPP loan). We find a much higher correlation between PPP loan volumes and state-level employment declines for loans issued in 2021 than those issued in 2020 (see online Appendix Figure B.1), suggesting that this targeting was more than nominal. Nevertheless these loans do not appear to have boosted employment, as we show below.

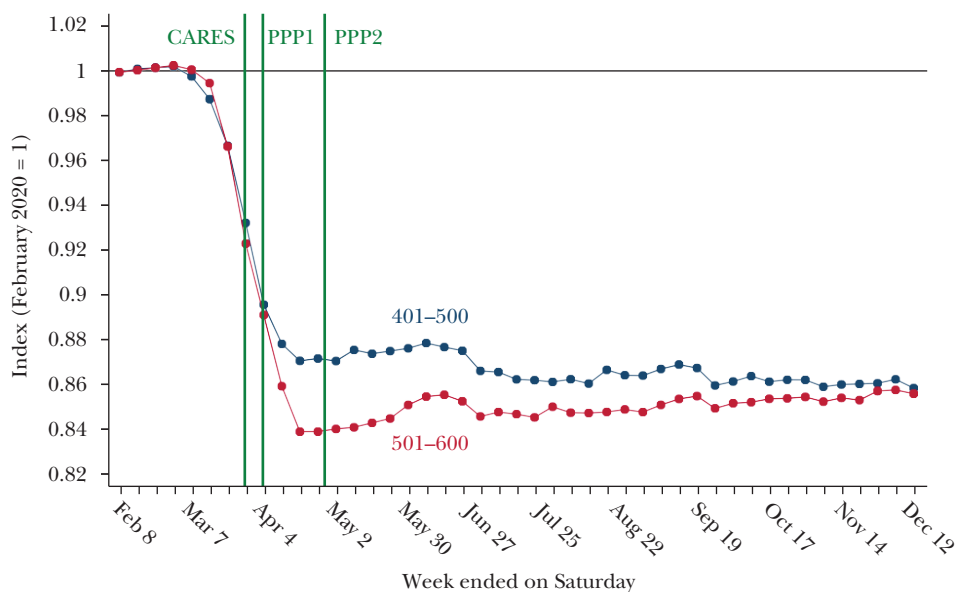
What Did the Paycheck Protection Program Accomplish?

Supporting Employment

A first step in calculating where the Paycheck Protection Program money went is to determine what fraction of funding went to paychecks that would otherwise not have been paid. Because PPP was ultimately taken up by almost all small businesses, we lack an ideal control group for making experimental comparisons. Nevertheless, a burgeoning literature, our own analysis included, indicates that PPP substantially boosted payroll employment.

The simplest and arguably most credible—though not necessarily most complete—method to assess the employment effects of the Paycheck Protection Plan is to compare the trajectory of employment at firms below the 500-employee initial-eligibility threshold to employment at ineligible firms above this threshold during the course of the pandemic. Figure 1—which is similar to our analysis in Autor et al. (2020)—presents this comparison using ADP payroll data. Employment is indexed to each firm's average level of employment in February 2020

Figure 1

Employment by Firm Size for Industries with PPP Eligibility at 500 Workers

Source: Author's calculations using ADP data.

Note: Each series represents average employment for firms with that particular range of workers during 2019 (on average) and in February 2020. Data are weighted by each firm's employment as of February 2020. Sample reflects firms that were present in the ADP data for all 12 months of 2019.

(immediately before the pandemic) for two employment size classes: 401–500 employees (in blue) and 501–600 employees (in red). Employment declines in parallel for these groups of firms at the start of the crisis. Following the launch of PPP, these trends diverge, with employment at firms that are likely eligible for PPP loans (401–500 employees) falling by substantially less than employment at firms that are likely ineligible (501–600). Approximately a month after the start of the PPP, employment had fallen by approximately 4 percent less at likely-eligible firms than at likely-ineligible firms. In the months thereafter, employment levels relative to baseline at likely-eligible and likely-ineligible firms gradually converged, with the difference falling to less than 2 percent by the start of July 2020. It disappeared altogether by September 2020.

Our formal econometric analysis of the employment effects of the Paycheck Protection Program in Autor et al. (2020) exploits this comparison of firms above versus below the size eligibility threshold, while additionally controlling for the differential impact of the pandemic across industries and states. After accounting for the fact that not all eligible firms received a loan, particularly in the initial months of the program, we estimate that taking out a PPP loan boosted firm employment by between 4 and 10 percent in mid-May and by 0 to 6 percent by the

end of the year.² Our best evidence is that about 2.97 million jobs per week were preserved by the Paycheck Protection Program in the second quarter of 2020, and 1.75 million jobs per week were preserved in the fourth quarter. Chetty et al. (2020) and Hubbard and Strain (2020) conduct similar analysis exploiting the eligibility size threshold, using non-ADP data sources, and reach broadly similar conclusions. Assuming that the employment effect declines linearly from its peak in May 2020 to zero by June 2021 implies that PPP saved 1.98 million worker years of employment at the very substantial cost of \$258,000 per worker-year retained.

These estimates based on eligibility thresholds are subject to an important caveat: because they focus on firms just above and below the 500 employee size-eligibility threshold for Paycheck Protection Program loans, they may not capture the effect of such loans on smaller firms. If smaller firms are more liquidity constrained and hence more likely to shrink or shut down during the pandemic (Chodorow-Reich et al. 2021), the threshold-based estimates will likely underestimate the effects of PPP at these firms and, by implication, understate the full effect of PPP.

To develop causal effect estimates that cover a broader set of treated firms, a number of papers exploit an event-study approach that compares employment at firms receiving a loan early in the program period to employment at firms receiving a loan later. This approach potentially captures the effect of PPP loans on small firms that are well below the eligibility threshold, though it comes at a cost of focusing only on the early months of the program, before most firms had taken loans.³

We complement existing event-study estimates using the vast ADP database, which offers substantial precision and a sample frame identical to that used for the size-threshold analysis above. To implement the event-study using the timing of loan takeup, we merge PPP loan-level data from the Small Business Administration to our sample of employers from ADP. This provides the precise date of PPP loan approval for each matched firm within a sizeable sample of firms with fewer than 50 employees.

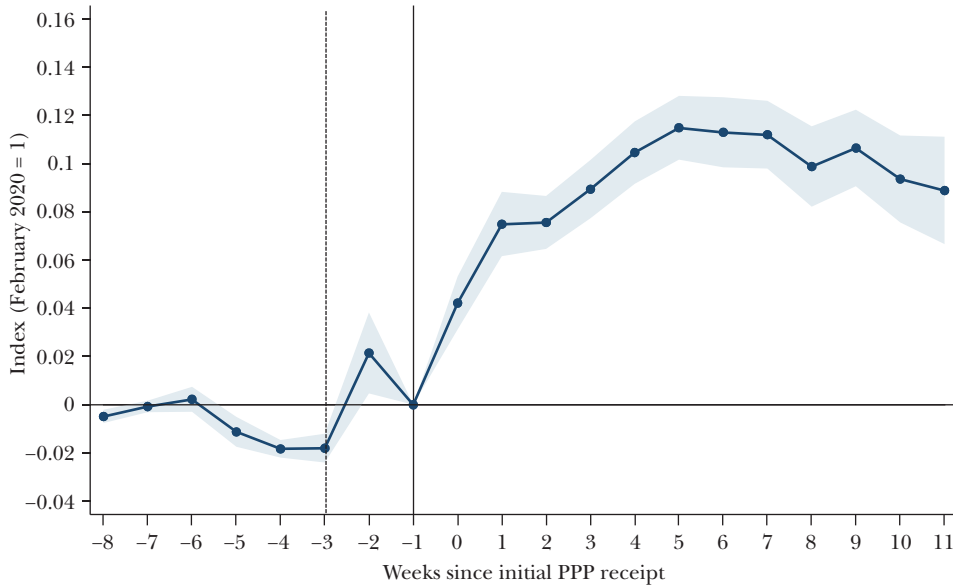
Figure 2 presents our timing-based estimates which trace out the effect of receipt of a Paycheck Protection Program loan on employment at firms with fewer than 50 employees.⁴ The employment trend prior to loan approval is roughly

²Adjusting for incomplete takeup means rescaling our Intent-to-Treat (ITT) estimates by the takeup rate to obtain Treatment-on-the-Treated estimates (TOT).

³Papers using the event-study approach obtain a range of employment effect estimates. The first to employ this approach, Granja et al. (2020), finds aggregate employment effects that are comparable to those found by the eligibility threshold papers. Estimates in Li and Strahan (2020) imply a much smaller boost to employment, however, while those in Bartik et al. (2021), Doniger and Kay (2021), Faulkender et al. (2020), and Kurmann et al. (2021) point toward a larger employment effect.

⁴A rapidly growing literature—for example, Goodman-Bacon (2021), Callaway and Sant’Anna (2020), Sun and Abraham (2020)—highlights the problems that arise in event-study estimates when the magnitude of the treatment effect is correlated with the timing of treatment. We resolve this issue using the approach proposed by Sun and Abraham (2020): we estimate and then average separate treatment effects for each of the first eleven cohorts of borrowers, where cohort refers to week of loan issuance, while using the final seven cohorts as a comparison group. We choose the final seven cohorts to ensure a sufficient sample size. Using only those firms receiving a PPP loan in the final week of the program yields qualitatively similar results.

Figure 2
Event-Study Employment Effects at Firms Sized 1–49



Source: Authors’ analysis of SBA and ADP data using Sun and Abraham (2020) “eventstudyinteract” STATA implementation.

Notes: Estimates from Sun and Abraham (2020) event-study interaction estimator on the sample of loan-matched ADP firms with between 1-49 employees where firm size is defined using the average size in February 2020. The outcome variable—firm-level employment—is indexed to equal 1 in February 2020. The estimates are weighted by each firm’s employment as of February 2020 and include controls for 3-digit industry-by-week and state-by-week fixed effects. Standard errors are clustered at the 3-digit industry. All points to the right of the solid line represent post-treatment periods. Alternatively, accounting for the biweekly pay schedule of most ADP employers, and the back-filling used to establish start dates, all periods to the right of the dashed line can be viewed as post-treatment. See online Appendix Section D.4 for more details.

flat and about equal to zero, but begins rising on loan approval. Five weeks later, employment is roughly 12 percent higher, where it remains through the close of

We estimate the following specification:

$$y_{it} = \alpha + \sum_{c \in T} \sum_{g=-8}^{11} (\beta_{c,g} * PPP_{g,it}) * D_c + \theta_{jt} + \theta_{st} + \epsilon_{it}$$

where y_{it} is total employment for firm i at week t indexed to equal 1 in February of 2020, θ_{jt} is a vector of NAICS 3-digit industry j -by-week t fixed effects, θ_{st} is a set of state s -by-week t fixed effects, and $PPP_{g,it}$ is a dummy variable equaling one if firm i at time t was approved for a PPP loan g weeks ago; $g = 0$ denotes the week of approval and the week prior to approval ($g = -1$) is the omitted category. D_c is a dummy variable denoting the week of PPP receipt for each cohort in the treatment set T (the first week through the eleventh week of the program). For additional details of how we implement the equation, see the online Appendix.

the outcome window. The relatively flat pre-trend centered around zero, and the sharp upward break after approval, are consistent with the interpretation that we are detecting a causal effect of PPP loans on small firm employment. We emphasize that these results indicate that small firms shrank *relatively* less after receiving a PPP loan as compared to firms not yet receiving a loan—not that their employment rose during the pandemic. The fact that the estimated effect on small firm employment is roughly *twice* as large as what we estimate for larger firms supports the view that smaller firms received a bigger employment boost from PPP.

Combining the results from Autor et al. (2020) for larger firms with the smaller firm results in Figure 2, we estimate that PPP loans originating in 2020 preserved about 3.0 million job-years at an average cost of \$169.300 per job-year saved. We use this result below when calculating the share of PPP funds that accrued to paychecks.

While our findings in Figures 1 and 2 capture the employment effects of loans issued in 2020 from the first two tranches of PPP funding, we know of no similar evidence on the consequences of the third major tranche of \$278 billion in PPP loans issued in 2021. To complete this picture, we estimate difference-in-difference threshold eligibility results analogous to those in Autor et al. (2020) for the second draw PPP loans which constituted the majority of third tranche loans issued in 2021 (comparing employment at firms above and below the 300 worker eligibility threshold for second draw loans).

Despite seemingly better targeting than the 2020 loans, we find no evidence in Figure 3 that the 2021 second-draw loans boosted employment, perhaps because they were issued too late to be relevant, after the economic recovery was well underway. If this interpretation is correct, it affirms that Congress was wise to prioritize speed over precision in dispatching the initial two tranches of PPP loans.

Preventing Firm Exits

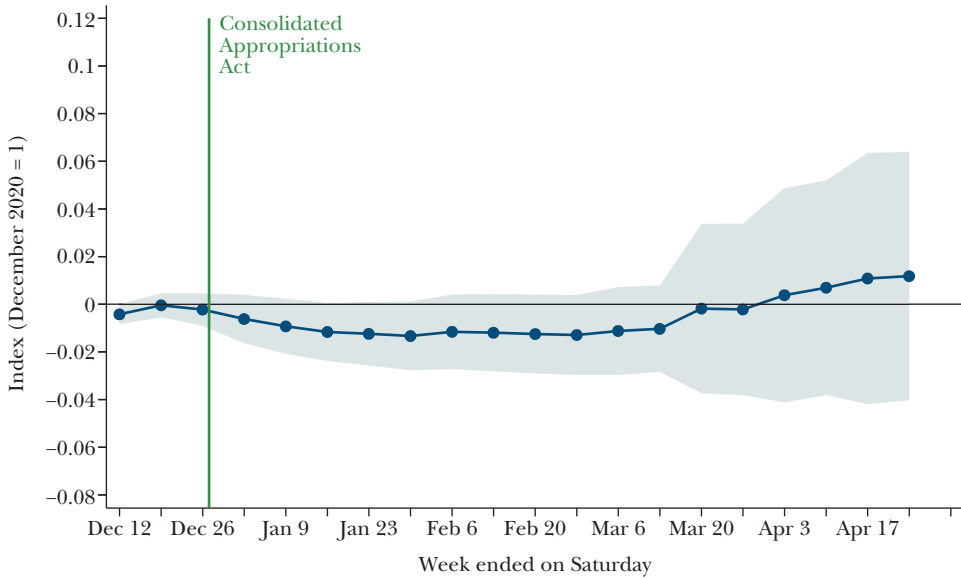
The spike in business closings during the COVID pandemic was historic. The Business Employment Dynamics database collected by the Bureau of Labor Statistics finds that employment at closing firms, which hovered at about 1 million worker per quarter for the last three decades, spiked to 2.1 million workers in the second quarter of 2020. Evidence in Crane et al. (2020) and Kurmann et al. (2021) corroborates these trends using a number of alternative indicators, including data from ADP.

A key justification for the Paycheck Protection Program was to prevent a contagion of business closures that would cause longer-term economic damage (Hubbard and Strain 2020). Business deaths—as distinct from business contractions and temporary closures—may potentially produce lasting economic harm not only by forcing the costly reallocation of physical capital, but also by permanently destroying worker-firm relationships and the associated match-specific capital (Farooq et al. 2020). Indeed, the prevalence of recall hires—as opposed to new hires—when firms rebound from contractions underscores the importance of match-specific capital to both employers and employees (for example, Fujita and Moscarini 2017).

We can observe the importance of firm closures for employment losses during the pandemic in the ADP data. Figure 4 groups firms into size classes based on their

Figure 3

Difference-in-Differences Eligibility Threshold Employment Effect, Second Draw Loans in 2021



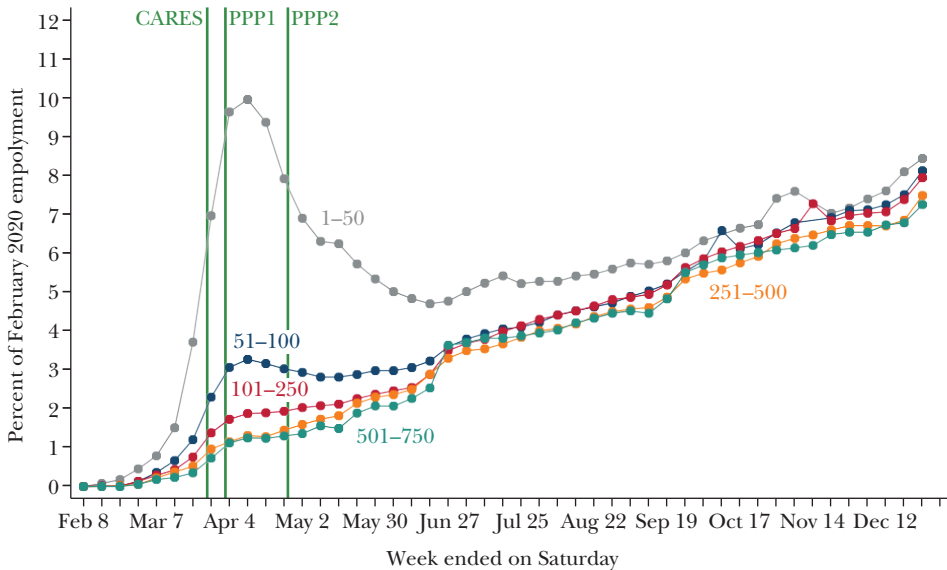
Source: Authors’ analysis of ADP data.

Note: Each firm’s size is determined using average employment in 2019 and employment in February 2020, the same size measure as used in Autor et al. (2020). Regressions are weighted by firm size as of February 2020 and include controls for state-by-week and industry-by-week effects. Standard errors are clustered at the 3-digit NAICS industry level. Sample comprises firms that were present in the ADP data for all 12 months of 2019. The specification is analogous to the difference-in-difference eligibility threshold specification in Autor et al. (2020); treatment and control groups are based on being within a range of 100 below and above, respectively, the 300 eligibility threshold for second draw loans. See the online Appendix for additional discussion.

February 2020 (pre-pandemic) employment and reports the share of their original employment that is lost due to firm shutdowns in each week between February and December 2020. Shutdowns are heavily concentrated among small firms: fully 10 percent of employment at firms that had 1–50 employees in February 2020 was lost due to shutdowns by early April 2020. For firms with more than 50 workers, these losses were only one-tenth to one-third as large. (We note that the general upward slope of the series in this figure is expected since some fraction of firms inevitably closes each year.⁵)

⁵We define a firm as shutdown if it has no paid employment in a given week. Although we cannot definitively determine whether firms that appear to be shutdown in the ADP data have shuttered business or rather stopped utilizing ADP’s payroll services, we expect that the *spike* in apparent shutdowns during the pandemic primarily reflects firms dropping to zero employment rather than discontinuing ADP’s services.

Figure 4

Share of February 2020 Employment at a Firm with Zero Employment

Source: Authors' analysis of ADP data.

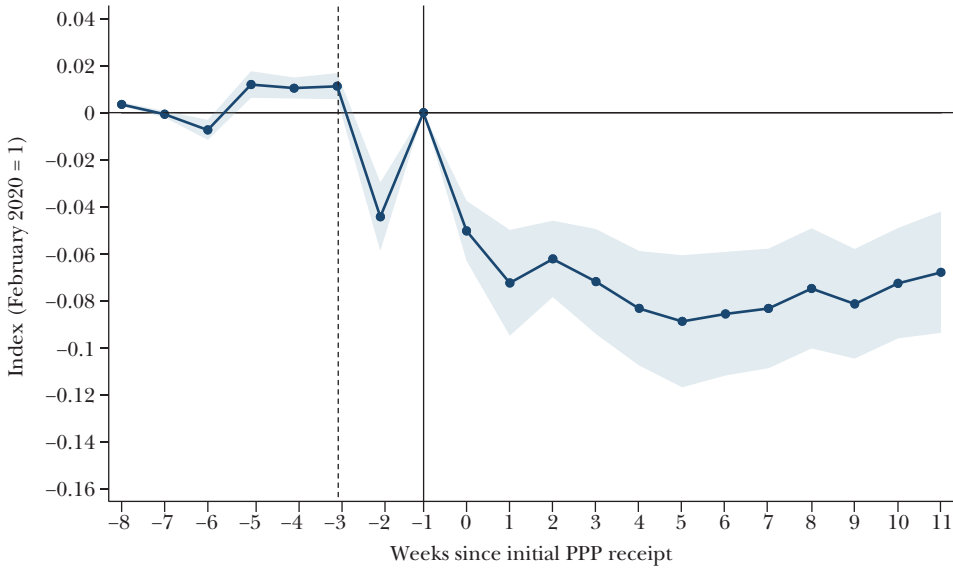
Note: Each series represents the size-weighted share of firms with zero employment in a given week, using February 2020 employment as the weight. Sample reflects firms that were present in the ADP data for all 12 months of 2019, and excludes firms in NAICS 72.

Figure 4 also offers tantalizing evidence that the Paycheck Protection Program may have inhibited firm closures or spurred reopenings. Among firms sized 1–50 and 51–100, firm shutdowns peaked shortly after PPP loans began flowing and rapidly reversed course thereafter. By June 2020, the fraction of employment at small firms lost due to closure was only half as large as in April—meaning that many had reopened.

Following recent work by Dalton (2021), we test whether the receipt of a Paycheck Protection Program loan affects the probability that firms with fewer than 50 employees remain open (or reopen after closure). Using event study estimates akin to those above for small-firm employment, we find in Figure 5 that PPP loans reduced employment losses due to small-firm closures by about 8 percentage points five weeks after loan receipt. Since our earlier results in Figure 2 found a peak PPP effect on small-firm employment of 12 percentage points at week five, we infer that about two-thirds of the employment-preserving effect of PPP loans on very small firm employment was due to PPP keeping the lights on at establishments that would have otherwise shuttered—at least temporarily.⁶

⁶One anomaly in our results in Figure 5 is that the estimated employment effects of PPP receipt at small firms appear to start a week too early relative to loan receipt. A possible explanation is that a large

Figure 5

Employment Change Due to Firm Closure, Firms Sized 1–49

Source: Authors' analysis of SBA and ADP data using Sun and Abraham (2020) "eventstudyinteract" STATA implementation.

Notes: Estimates from Sun and Abraham (2020) event-study estimator on the sample of loan-matched ADP firms with between 1–49 employees where firm size is defined using the average size in February 2020. The analysis is implemented by setting y_{it} in the equation in footnote 4 equal to one if firm i has zero employment in week t and zero if it has positive employment. The specification includes controls for 3-digit industry-by-week and state-by-week fixed effects. Standard errors are clustered at the 3-digit industry. The coefficient vector displayed in the figure can be interpreted as the percentage point effect of the PPP on employment due to firm closure as the specification is weighted by pre-pandemic firm employment. All points to the right of the solid line represent post-treatment periods. Alternatively, accounting for the biweekly pay schedule of most ADP employers, and the back-filling used to establish start dates, all periods to the right of the dashed line can be viewed as post-treatment. See online appendix Section D.4 for more details.

Ultimately, permanent business closure proved less pervasive than many had anticipated at the pandemic's onset. The Paycheck Protection Program may be part of the reason. Because our methodology permits examining firm closures only over the short run, we cannot assess whether PPP averted permanent firm exits or mainly temporary closures. Using a related methodology, Dalton (2021) finds that the

fraction of ADP paycheck recipients are paid biweekly, and this payment scheme blurs the observable timing of any discrete event over the prior 13 days. Concretely, imagine that a firm's two-week pay period begins July 17 and ends July 31. After receiving a PPP loan, that firm reopens its doors on July 30th. In our estimation, this firm will show an employment jump on July 17, even though all of its hires occurred 13 days later, when the loan was issued.

PPP effect on small-firm closures waned somewhat over the ensuing seven months, indicating that some of the PPP effect on closure was temporary, not permanent, in nature. For larger firms around the 500 employee eligibility threshold, we find no consistent evidence that the PPP influenced shutdowns, either over the short- or longer-term (for details, see online Appendix figure E.1). Despite bolstering jobs during the pandemic, PPP may not have had a pronounced effect on preserving intangible business capital. More work is needed to definitively assess the effect of the PPP on permanent business closure.⁷

Reducing Commercial Delinquency

Alongside preserving jobs and keeping firms open during the pandemic, the Paycheck Protection Program may have indirectly benefited creditors of small businesses—landlords, banks, holders of mortgage-backed securities, and suppliers—by keeping payments flowing. There is limited evidence on the effect of PPP on loan recipients' ability to pay creditors, but the evidence that exists suggests the impact was positive. Exploiting differences in the tendency of small versus large commercial properties to have PPP-eligible tenants, Agarwal et al. (2021) find that PPP significantly blunted the rise in commercial mortgage delinquency rates during the pandemic, particularly in the retail sector. Using survey data and a variant of the event-study strategy discussed above, Granja et al. (2020) also find that PPP decreased delinquency on mortgages and other payments.

Where Did the Money Go?

The estimates above provide a key input for answering our motivating question—that is, where did the money go? Using the employment effects estimated above, along with many other data sources, we estimate the incidence of the \$510 billion in Paycheck Protection Program loans issued in 2020 across the household income distribution. We further compare this incidence to other pandemic economic assistance programs. Additional information on these calculations can be found in section F of the online Appendix.

Proximate Recipients: Workers versus Non-workers

Paycheck Protection Program funds were paid to businesses. In turn, businesses used these funds to pay three proximate groups of beneficiaries: workers who otherwise would have been laid off; creditors and suppliers who otherwise would not have been paid; and owners and shareholders of PPP-receiving firms as residual claimants in cases where businesses would have met some or all of their payroll and other financial obligations absent PPP (also known as “windfall profits”). The distribution

⁷Other recent work provides mixed evidence on these outcomes. Granja et al. (2020) find little evidence of a PPP effect on firm shutdown. Bartik et al. (2021) and Kurmann et al. (2021), though, find that PPP mitigated business shutdowns.

of PPP funds among these groups—workers versus non-workers, in particular—matters for our accounting exercise because different groups represent different parts of the household income distribution.

We focus first on payments to workers. As documented above, Paycheck Protection Program loans issued in 2020 modestly raised employment at recipient firms. To convert these employment effects into payroll expenditures, we use the main estimates reported above from Autor et al. (2020), who find that PPP boosted employment by about 6 percent in mid-May 2020, with effects tapering off gradually thereafter. These numbers imply that PPP preserved about 2.97 million jobs per week in the second quarter of 2020 and about 1.75 million jobs per week by the fourth quarter of 2020. Assuming a linear trend decline in this program effect, PPP would have had zero employment effects by June 2021. Converting these weekly job numbers into job-years (that is, one worker for one year), implies that PPP preserved about 1.98 million job-years of employment at a cost of \$258,000 per job-year saved (that is, \$510 billion/1.98 million). We assume that actual employee compensation for each saved job averaged \$58,200 since the average weekly wage from the Current Population Survey in February 2020 is \$786 (truncating at an annual wage of \$100,000 above which the PPP did not provide additional support per worker) and, on average, total compensation is 42 percent larger than wages according to BLS Employer Costs for Employee Compensation data ($\$786 \times 52 \times 1.42$). The 1.98 million job-years saved then imply that \$115 billion in PPP loans ($\$58,200 \times 1.98$ million) accrued to employee paychecks.

We also produce an alternative estimate of the amount of Paycheck Protection Program loans accruing to compensation based on our 3.0 million job-years saved estimate which combines the results from Autor et al. (2020) and the larger effects for smaller firms in Figure 2. Continuing to assume that compensation at retained jobs averaged \$58,200 implies that \$175 billion in PPP compensation went to paychecks. It is likely that this \$175 billion estimate is an upper bound on the share of PPP funds flowing to worker compensation. Some event-study estimates for the *entire* size distribution of PPP-eligible firms, including small firms, find an overall peak PPP employment effect of approximately 6 to 8 percent (for example, see online Appendix Figure D.1 and Dalton 2021). These estimates are more in line with our smaller \$115 billion estimate. Additionally, our assumption of a smooth trend decline in PPP's impacts through June 2021 is generous.⁸ Moreover, we are not accounting for loans issued in 2021 where our evidence suggests the PPP failed to boost employment; doing so would further lower the estimated share of PPP loans that flowed to workers relative to non-workers and hence substantially raise the estimated cost per job-year preserved.

These bounds of \$115 billion to \$175 billion in Paycheck Protection Program funds accruing directly to paychecks imply that between 23 percent and 34 percent

⁸In Autor et al. (2020), we detect no statistically significant impact of PPP on employment after July 2020. Because the point estimates remain non-zero through December 2020, we extrapolate the entire series out until it is numerically zero in June 2021.

of the first two tranches of PPP dollars totaling \$510 billion supported jobs that would otherwise have been lost. By implication, the remaining \$335 to \$395 billion (66 to 77 percent) accrued to owners of business and corporate stakeholders, including creditors and suppliers, and others.

The Household Distributional Incidence

To trace the flow of Paycheck Protection Program payments from their proximate recipients to their household incidence requires information on the income distributions of both worker and non-worker (that is, owner) beneficiaries. Starting with the worker beneficiaries, we estimated above that, at the high end, \$175 billion in PPP money flowed to workers whose jobs were saved. We assume that the distributional incidence of those funds followed the distribution of job loss in 2020 by household income quintile. To estimate this distribution, we first measure employment declines across the weekly wage distribution using the Current Population Survey Outgoing Rotation Group (CPSORG) files. Pandemic job losses were largest for low-paid workers: total employment fell by 17.8 percent from March 2020 through the end of 2020 among the lowest-paid (first) quintile of workers; by 10.6 percent in the second quintile; by 6.0 and 2.2 percent in the third and fourth quintiles, respectively; and by a substantial 8.7 percent among the highest quintile of earners. We convert these job loss percentages into average weekly wage losses by multiplying each by the February 2020 pre-COVID average weekly wage within quintile. From there, it is straightforward to calculate the share of compensation lost by weekly wage quintile, which we impute to the household income distribution using March CPS data on the joint distribution of weekly wages and household income.⁹

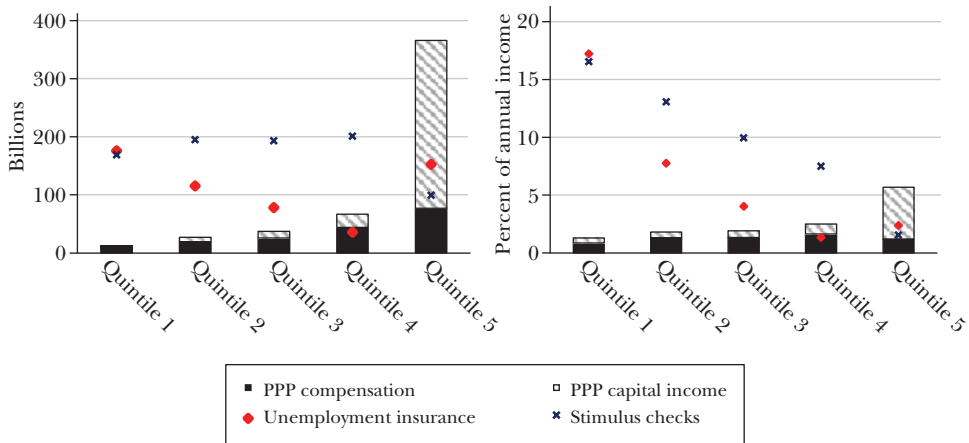
We make an analogous (but simpler) imputation for the household incidence of the \$335 billion in Paycheck Protection Program fund payments that flowed to non-worker beneficiaries, such as creditors and suppliers who otherwise would not have been paid and owners and shareholders of PPP-receiving firms. Specifically, we use the Congressional Budget Office's (2020a) most recent estimates on the distribution of capital incomes by type to distribute the funds across households. We do not attempt to account for the flow of PPP funds from proximate and subsequent recipients—for example, a PPP-receiving firm's supplier pays its workers or a worker at a PPP-receiving firm pays her landlord. Thus, our exercise is in the spirit of the static distributional incidence analyses performed for tax policies by the Joint Committee on Taxation.

Unlike Paycheck Protection Program payments, which went to businesses, transfer payments made by the two other major federal pandemic emergency assistance programs—pandemic unemployment insurance payments and household

⁹Specifically, we calculate $S_q \equiv T \times \frac{U_q \times W_q}{\sum_{q=1}^5 U_q \times W_q}$, where T is total PPP dollars that support employment, U_q is quintile q 's share of job losses during the pandemic, and W_q is quintile q 's wage in February 2020 prior to the pandemic. We then map from the *weekly wage* distribution to the *household income* distribution using March CPS supplement data on the probability that a worker in a given weekly wage quintile is in each household income quintile.

Figure 6

Distribution of Paycheck Protection Program, Unemployment Insurance, and Stimulus Check Payments



Source: Authors' analysis of CBO, Census Bureau, BEA, BLS ECEC, and Current Population Survey microdata, and estimates from Autor et al. (2020), Bhutta et al. (2020), and Boesch et al. (2021).

Note: See online Appendix for details of calculations. Calculations are based on our upper-end (most generous) estimate of the wage compensation supported by PPP. Online Appendix Figure F.1 reports the same calculations using our lower-end estimate of the wage compensation supported by PPP.

stimulus payments—went directly to households and workers. The size of these payments rivaled those of PPP, as noted above. To facilitate comparison with PPP, we calculate the distributional incidence of these as well.

For household payments, we use incidence data from Bhutta et al. (2020), who analyzed the effect of stimulus payments on household finances using the Survey of Consumer Finances. For unemployment insurance, we calculate approximate shares of benefits paid during the pandemic—including regular state programs and the pandemic enhancements to unemployment insurance—using, as a starting point, our estimates of average wages lost during the pandemic, and applying the methodology above for apportioning the PPP funds to paychecks. We combine these data on wages lost by quintile with simple estimates of the unemployment insurance replacement rate by quintile, which we estimate using same CPS ORG data. This calculation accounts for both pandemic supplements to weekly unemployment insurance benefits and the portion of Pandemic Unemployment Assistance payments that went to the self-employed, as estimated by Boesch et al. (2021).

Panel A of Figure 6 reports the distribution of Paycheck Protection Program, unemployment insurance, and household payments in billions of dollars across household quintiles. As the figure makes clear, the distribution of PPP loans overwhelmingly accrued to high-income households. Of the \$510 billion in PPP loans provided in 2020, we estimate that only \$13.2 billion ultimately flowed to households in the bottom fifth of the income distribution, and that \$130.8 billion flowed to the second through fourth quintiles. The remaining \$365.9 billion (72 percent)

flowed to the top fifth of household income. This skew reflects two features of PPP. First, high-wage earners are found in high-income households. Though the PPP offered loans to support only up to \$100,000 in annual earnings, even with this truncation, the top fifth of households accounts for about 35 percent of wage and salary earnings. Second, the distribution of capital ownership is even more right-skewed than the distribution of wage earnings—with the top fifth of households commanding 86.2 percent of capital income—meaning that subsidies to businesses are ultimately subsidies to high-income households.

If we additionally account for the \$100 billion in tax credits that Congress granted to PPP-receiving businesses in January 2021 by making PPP payments nontaxable, we conclude that an additional \$85 billion flowed to the top quintile of households, raising that quintile's total PPP distribution further from \$365.9 to \$450.9 billion.¹⁰

In comparison, both household stimulus payments and pandemic unemployment insurance payments were far less regressive than in the Paycheck Protection Program. The incidence of household stimulus checks in dollar terms was close to uniform across the lower four income quintiles. Moreover, due to the income caps that Congress set on household payments, the incidence of these payments was much smaller for the highest quintile of households.

Meanwhile, the incidence of unemployment insurance during the pandemic was weighted towards both the upper and lower tails of the household income distribution. We estimate that 31.5 percent (\$175.6 billion) and 20.7 percent (\$115.4 billion) in pandemic unemployment insurance payments went to the bottom fifth and second-to-bottom fifth of households, respectively (red diamonds in panel A of Figure 6). Surprisingly, the top fifth of households received a bit more than one-quarter of unemployment insurance benefits. This occurred both because the highest income quintile of wage and salary workers sustained substantial employment losses during the pandemic (as documented above), and because the Pandemic Unemployment Assistance (PUA) program allowed self-employed business owners—who tend to have high incomes—to collect unemployment insurance benefits. Estimates from Boesch et al. (2021) suggest that self-employed business owners received about 40 percent of Pandemic Unemployment Assistance insurance benefits.

Panel B of Figure 6 recasts these distributional incidence figures into household annual income replacement rates rather than dollar transfers. Both stimulus checks and unemployment insurance payments replaced about 17 percent of the incomes of the lowest quintile of households, with much lower shares at higher quintiles. Thus, although the combination of these three programs is highly regressive

¹⁰ We make this calculation by attributing the \$100 billion in forgone tax revenues to the share of business income that goes to the top fifth of households, which is approximately 85 percent. An alternative calculation yields a comparable conclusion: the Congressional Budget Office estimates that the average marginal tax rate on capital income is 18.6 percent; noting that PPP distributed \$510 billion in 2020 and that the top fifth of households commands about 85 percent of capital income, we obtain a tax benefit for the top fifth of \$81 billion.

in dollar terms, it is roughly progressive in replacement rate terms due to the highly skewed distribution of US household incomes.

Macroeconomic Benefits

An additional benefit of these transfers programs is that they provided stimulus during a time of rapid economic contraction. The short-term macroeconomic boost of a program during a recessionary period is conventionally linked to the marginal propensity to consume (MPC) of those who receive benefits from the program. Cashin et al. (2018) provide estimates of the MPC for different types of fiscal shocks based on characteristics such as: the type of policy change (say, tax versus transfer payment); who is receiving the benefit (say, low-income households versus corporations); and whether the flow of benefits is temporary or permanent. These MPC estimates are informed by a publicly available macroeconomic model, FRB/US, used by the Federal Reserve Board staff (described in Brayton et al. 2014), and by the relevant empirical literature.

Using these marginal propensity to consume estimates, we offer a back-of-the-envelope comparison of the degree of stimulus provided by the three main programs mentioned above: the Payment Protection Program, stimulus payments, and pandemic-enhanced unemployment insurance. This calculation relies on the following assumptions:

1. Since unemployment insurance payments are generally made to households that are highly liquidity-constrained, the marginal propensity to consume out of unemployment insurance payments is one (Cashin et al. 2018).
2. Because stimulus payments are made to a broad mixture of households across the income distribution, we use the estimate from Cashin et al. (2018) for the MPC of general, temporary transfers to households of 0.5.
3. The part of the Paycheck Protection Program that flows through to wages is similar to unemployment insurance, and thus has a plausible marginal propensity to consume of one.
4. For the part of PPP that flows to non-workers, we use the estimates in Cashin et al. (2018) for the MPC of temporary corporate tax cuts of 0.2; this relatively low marginal propensity to consume is consistent with these funds flowing disproportionately to the upper quintile of the income distribution.

Weighting these last two components together, we obtain an overall marginal propensity to consume out of PPP loans of about 0.5, which is comparable to stimulus checks (where we have an imputed MPC of 0.5) and much lower than unemployment insurance payments (where we have imputed an MPC of one).¹¹

¹¹We noted in the previous section that a substantial share of Pandemic Unemployment Assistance recipients were likely high income self-employed business owners who might be expected to have a lower marginal propensity to consume than the one we assume here for unemployment insurance recipients.

This illustrative calculation thus suggests the PPP loans and stimulus checks were roughly equally effective at boosting spending, and both were much less effective on this margin than pandemic unemployment insurance.

These estimates have the virtue of transparency. They also have shortcomings. The pandemic environment surely generated non-normal household and business behavior. Extraordinarily high replacement rates delivered by enhanced unemployment benefits may have diminished the marginal propensity to consume of recipients. The substantial share of payroll income received by the top quintile from the Paycheck Protection Program also suggests that treating this income as similar to unemployment insurance probably overstates the MPC. Finally, these estimates quantify only the transfer's initial boost to aggregate demand; they do not capture aggregate supply effects, such as that arising from preventing firm bankruptcies, or subsequent general equilibrium effects. Fortunately, the Congressional Budget Office (2020b) has also estimated the boost to GDP per dollar for these same pandemic programs, carefully accounting for the pandemic environment and the specifics of each program. CBO also strives to capture the full, general equilibrium effect of each program, including its potential impact on business closure. CBO concludes that the enhanced unemployment and stimulus checks were far more effective at boosting GDP than was PPP. Specifically, the CBO estimates a per dollar boost to GDP of 0.36 for the PPP and 0.60 and 0.67 for stimulus checks and enhanced unemployment insurance benefits, respectively. Taking account of the highly distributionally-skewed incidence of PPP payments, we concur that PPP was likely the least effective of the three programs in boosting the macroeconomy.

Lessons Learned from the Paycheck Protection Program Experience

The US small business sector appeared at risk of collapse at the outset of the pandemic. To avert this collapse, Congress enacted the Paycheck Protection Program, which successfully distributed vast amounts of aid to the near-universe of eligible small businesses in the space of a few months. Our best evidence to date indicates that the PPP's economic impacts were less than hoped: it preserved only a moderate number of jobs at a high cost per job-year retained and transferred resources overwhelmingly to the highest quintile of households.

These outcomes should not, however, be viewed first and foremost as programmatic failures. The PPP's regressive distributional incidence and its limited efficacy as economic stimulus stem from the program's absence of targeting. This absence, in turn, reflected necessity. Given the time constraints and, more profoundly, the

Nonetheless, even if we assumed an MPC of *zero* for self-employed PUA recipients, the overall MPC out of unemployment insurance would be about 0.7 (since non-PUA benefits were about 70 percent of total unemployment insurance), which is still higher than our estimates of the MPC out of PPP and stimulus checks.

lack of existing administrative infrastructure for overseeing targeted federal support to the entire population of US small businesses at the onset of the pandemic, we strongly suspect that Congress could not have better targeted the Paycheck Protection Program without substantially slowing its delivery. We thus concur with Bartik et al. (2021) that policymakers made a defensible trade-off between speed and targeting in the PPP's design.

However, if the Paycheck Protection Program was a logical answer to a highly constrained question, a forward-looking lesson from the PPP experience is that the United States should invest now to relax those constraints. We have emphasized that the PPP had dual goals: preserving jobs and providing liquidity. These goals could be better served with the administrative capacity to address these issues directly and separately, thus enabling better targeting and a more progressive incidence. The primary job retention goal of the Paycheck Protection Program could in the future be better achieved through an expanded program to encourage "work-sharing," which refers to a policy in which employers, when faced with an economic downturn, are encouraged to reduce hours worked more broadly across the workforce rather than laying off a narrower group outright. In effect, the government program ends up paying partial unemployment to many, rather than full unemployment to some.

Currently, 26 US states have a work-sharing program through their unemployment insurance systems, but these were not well-subscribed where available during the COVID recession. A number of proposals over the last decade that advocate for expanded work-sharing suggest that to reach broader coverage, such programs should be simplified and automated (Abraham and Houseman 2014; Strain 2020; Dube 2021). A work-sharing program can target firms of all sizes that are cutting hours or employment, not just small firms. Additionally, with sufficient administrative capacity developed in normal times, the progressivity of the program could be altered as policymakers deem appropriate.

A separate liquidity provision program could then be targeted primarily at small firms, which are more likely to be liquidity-constrained. Moreover, with better information systems operational, liquidity could be provided in proportion to firms' decline in revenues as well as firms' actual fixed obligations.

Distinct from the United States, many other high-income countries responded to the pandemic with a mixture of job retention incentives, including 1) work-sharing programs that allowed either partial or complete furloughs; 2) newly introduced wage subsidy programs, similar in many ways to the Paycheck Protection Program, that provided businesses with direct support for at least some fraction of their wage bill (OECD 2021). Both work-sharing and wage subsidy programs were targeted. Wage subsidy programs were explicitly targeted to firms that had experienced declines in revenue: for example, Canada's Employer Wage Subsidy was available to firms that experienced a year-over-year revenue drop of 30 percent (reduced to 10 percent later). In some countries, firms were entitled to wage subsidies on a sliding scale in proportion to their declines in revenues. By contrast, work-sharing programs were not *explicitly* targeted to distressed firms. But the requirement that firms reduce workers' hours to obtain assistance generally makes firm participation

unattractive absent a negative shock (for additional discussion in this symposium, see the paper by Giupponi, Landais, and Lapeyre). Of course, the details of these programs, such as the length of benefits and the extent to which non-payroll expenses are covered, as well as their efficacy, varies across countries.

A key lesson from these cross-national comparisons is that targeted business support systems were feasible and rapidly scalable in other high-income countries because administrative systems for monitoring worker hours and topping up paychecks were already in place prior to the pandemic. Lacking such systems, the United States chose to administer emergency aid using a fire hose rather than a fire extinguisher, with the predictable consequence that virtually the entire small business sector was doused with money. This approach may have been necessary, but it was desirable only because the United States lacked viable alternatives. By building administrative capacity in the years ahead, the United States could more deftly target, calibrate, and deploy its emergency business response systems when most needed. This investment will pay off in the near term in improvements of existing programs, such as work sharing—and will pay off again when the next crisis arises.

■ *At the Federal Reserve Board, we thank Christopher Kurz and Norman Morin for support, Kendra Robbins and Eleanor Warren for excellent research assistance, and Kevin Moore for updating the analysis in Bhutta et al. (2020). We thank Tolga Tuncoglu of ADP for superb assistance with matching the SBA PPP loan data into the ADP data. We thank Michael Dalton, Andrew Goodman-Bacon, Erik Hurst, Katie Lim, Adam Looney, Joseph Nichols, Ryan Nunn, Matthew Shapiro, Liyang Sun, and Eric Zwick for helpful discussion. Autor acknowledges financial support from the Smith Richardson Foundation (#20202252), Accenture LLP (#027843-0001), the Andrew Carnegie Fellowship (G-F-19-56882), and the Washington Center for Equitable Growth (APP-01666). The analysis and conclusions set forth here are those of the authors and do not indicate concurrence by other members of the Federal Reserve Board research staff, by the Board of Governors, or by ADP (ADP's data privacy policy can be found at <https://www.adp.com/about-adp/data-privacy.aspx>).*

References

- Abraham, Katherine G., and Susan N. Houseman.** 2014. "Proposal 12: Work Sharing to Reduce Unemployment." *Policies to Address Poverty in America*, edited by Melissa S. Kearney and Benjamin H. Harris, 1–10. Washington, DC: The Brookings Institution.
- Agarwal, Sumit, Brent W. Ambrose, Luis A. Lopez, and Xue Xiao.** 2021. "Did the Paycheck Protection Program Help Small Businesses? Evidence from Commercial Mortgage-backed Securities." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3674960.
- Amiram, Dan, and Daniel Rabetti.** 2020. "The Relevance of Relationship Lending in Times of Crisis."

- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3701587.
- Anbil, Sriya, Mark Carlson, and Mary-Frances Styczynski.** 2021. "The Effect of the PPPLF on PPP Lending by Commercial Banks." FEDS Working Paper 2021-030.
- Autor, David, David Cho, Leland Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz.** 2020. "An Evaluation of the Paycheck Protection Program Using Administrative Payroll Microdata." <https://economics.mit.edu/files/20094>.
- Bartik, Alexander W., Marianne Bertrand, Feng Lin, Jesse Rothstein, and Matthew Unrath.** 2020a. "Measuring the Labor Market at the Onset of the COVID-19 Crisis." *Brookings Papers on Economic Activity*.
- Bartik, Alexander W., Marianne Bertrand, Zoe Cullen, Edward L. Glaeser, Michael Luca, and Christopher Stanton.** 2020b. "The Impact of COVID-19 on Small Business Outcomes and Expectations." *Proceedings of the National Academy of Sciences* 117 (30): 17656–17666.
- Bartik, Alexander W., Zoe B. Cullen, Edward L. Glaeser, Michael Luca, Christopher T. Stanton, and Adi Sunderam.** 2021. "The Targeting and Impact of Paycheck Protection Program Loans to Small Businesses." NBER Working Paper 27623.
- Bernstein, Jared, and Jesse Rothstein.** 2020. "A Fast, Simple Way to Get Support to Workers without Paid Leave." *Washington Post*, March 10. <https://www.washingtonpost.com/outlook/2020/03/10/fast-simple-way-get-support-workers-without-paid-leave/>.
- Bhutta, Neil, Jacqueline Blair, Lisa Dettling, and Kevin Moore.** 2020. "COVID-19, The CARES act, and Families' Financial Security." *National Tax Journal* 73 (3): 645–72.
- Boesch, Tyler, Katherine Lim, and Ryan Nunn.** 2021. "What Did and Didn't Work in Unemployment Insurance during the Pandemic." *Federal Reserve Bank of Minneapolis*.
- Brayton, Flint, Thomas Laubach, and David Reifschneider.** 2014. "The FRB/US Model: A Tool for Macroeconomic Policy Analysis." <https://www.federalreserve.gov/econresdata/notes/feds-notes/2014/a-tool-for-macroeconomic-policy-analysis.html>.
- Brevoort, Kenneth P., John A. Holmes, and John D. Wolken.** 2010. "Distance Still Matters: The Information Revolution in Small Business Lending and the Persistent Role of Location, 1993–2003." FEDS Working Paper 2010-08.
- Callaway, Brantly, and Pedro H.C. Sant'Anna.** 2020. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–30.
- Cashin, David, Jamie Lenney, Byron Lutz, and William Peterman.** 2018. "Fiscal Policy and Aggregate Demand in the USA before, during, and following the Great Recession." *International Tax and Public Finance* 25: 1519–58.
- Congressional Budget Office.** 2020a. *The Distribution of Household Income, 2017*. Congressional Budget Office.
- Congressional Budget Office.** 2020b. "The Effects of Pandemic-Related Legislation on Output." Congressional Budget Office.
- Chetty, Raj, John N. Friedman, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team.** 2020. "How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data." NBER Working Paper 2743.
- Chodorow-Reich, Gabriel, Olivier Darmouni, Stephan Luck, and Matthew Plosser.** 2021. "Bank Liquidity Provision across the Firm Size Distribution." *Journal of Financial Economics*.
- Cororaton, Anna, and Samuel Rosen.** 2021. "Public Firm Borrowers of the US Paycheck Protection Program." *Review of Corporate Finance Studies* 10 (4): 641–93.
- Crane, Leland D., Ryan A. Decker, Aaron B. Flaen, Adrian Hamins-Puertolas, and Christopher J. Kurz.** 2020. "Business Exit During the COVID-19 Pandemic: Non-Traditional Measures in Historical Context." Finance and Economics Discussion Series 2020-089.
- Center for Economic Policy Research.** 2021. "CPS ORG Uniform Extracts, Version 2.5." Washington, DC.
- CRFB.** 2021. "Covid Money Tracker." <https://www.covidmoneytracker.org/>.
- Dalton, Michael.** 2021. "Putting the Paycheck Protection Program into Perspective: An Analysis Using Administrative and Survey Data." Bureau of Labor Statistics Working Paper 542.
- Decker, Ryan A., Robert J. Kurtzman, Byron F. Lutz, and Christopher J. Nekarda.** 2021. "Across the Universe: Policy Support for Employment and Revenue in the Pandemic Recession." *AEA Papers and Proceedings* 111 (May): 267–71.
- Doniger, Cynthia L., and Benjamin Kay.** 2021. "Ten Days Late and Billions of Dollars Short: The Employment Effects of Delays in Paycheck Protection Program Financing." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3747223.

- Dube, Arindrajit.** 2021. "A Plan to Reform the Unemployment Insurance System in the United States." Hamilton Project Policy Proposal 2021-03.
- Elmendorf, Douglas W., and Jason Furman.** 2008. "If, When, How: A Primer on Fiscal Stimulus." Hamilton Project Strategy Paper January 2008.
- Erel, Isil, and Jack Liebersohn.** 2020. "Does FinTech Substitute for Banks? Evidence from the Paycheck Protection Program." NBER Working Paper 27659.
- Farooq, Ammar, Adriana D. Kugler, and Umberto Muratori.** 2020. "Do Unemployment Insurance Benefits Improve Match Quality? Evidence from Recent U.S. Recessions." NBER Working Paper 27574.
- Faulkender, Michael, Robert Jackman, and Stephen I. Miran.** 2020. "The Job-Preservation Effects of Paycheck Protection Program Loans." US Treasury, Office of Economic Policy Working Paper 2020-01.
- Ilood, Sarah, Miriam King, Renae Rodgers, Steven J. Ruggles, Robert Warren, and Michael Westberry.** 2021. "Integrated Public Use Microdata Series, Current Population Survey: Version 9.0, Annual Social and Economic Supplement 2020." Minneapolis, MN: IPUMS.
- Fujita, Shigeru, and Giuseppe Moscarini.** 2017. "Recall and Unemployment." *American Economic Review* 107 (12): 3875–3916.
- Goodman-Bacon, Andrew.** 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.
- Granja, João, Christos Makridis, Constantine Yannelis, and Eric Zwick.** 2020. "Did the Paycheck Protection Program Hit the Target?" NBER Working Paper 27095.
- Harney, John, and Gregory Mott.** 2021. "IRS Blesses Tax Breaks on Forgiven PPP Loans after Law Change." *Bloomberg Tax*, January 6. <https://news.bloombergtax.com/daily-tax-report/irs-blesses-tax-breaks-on-forgiven-ppp-loans-after-law-change>.
- Hubbard, Glenn, and Michael R. Strain.** 2020. "Has the Paycheck Protection Program Succeeded?" *Brookings Papers on Economic Activity*.
- Joaquim, Gustavo, and Felipe Netto.** 2021. "Bank Incentives and the Impact of the Paycheck Protection Program." Unpublished.
- Kurmann, Andre, Etienne Lale, and Lien Ta.** 2021. "The Impact of COVID-19 on Small Business Dynamics and Employment: Real-Time Estimates with Homebase Data." Unpublished.
- Li, Lei, and Philip Strahan.** 2020. "Who Supplies PPP Loans (And Does it Matter)? Banks, Relationships and the COVID Crisis." NBER Working Paper 28286.
- OECD.** 2021. *OECD Employment Outlook 2021*. <https://www.oecd.org/employment-outlook/#Read-the-report>.
- Small Business Administration (SBA).** 2019. "2019 Small Business Profile." Technical Report, Office of Advocacy 2019. <https://cdn.advocacy.sba.gov/wp-content/uploads/2019/04/23142719/2019-Small-Business-Profiles-US.pdf>.
- Small Business Administration (SBA).** 2021. "Forgiveness Platform Lender Submission Metrics, November 28, 2021." https://www.sba.gov/document/report-ppp-forgiveness-platform-lender-submission-metrics-reports?utm_medium=email&utm_source=govdelivery (accessed November 30, 2021).
- Strain, Michael.** 2020. "Let's Provide Unemployment Benefits Without the Layoffs." *Bloomberg Opinion*, May 7.
- Sun, Liyang, and Sarah Abraham.** 2020. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (22): 175–99.

American Enslavement and the Recovery of Black Economic History

Trevon D. Logan

Every few decades, America “discovers” its history of enslavement. These rediscoveries are always about a contemporary aspect of race relations in the United States. In the 1890s, it was based on a reassessment of the Civil War and Reconstruction, which led to a valorization of the Confederacy and the formal establishment of Jim Crow laws that legalized discrimination and segregation of Black Americans. In the 1910s and 1920s, the rediscovery was an extension of a violent wave of race riots throughout the country, designed to keep Black Americans in their second-class political and economic place. In the 1950s, the rediscovery overturned many of the revisionist assumptions of the 1890s and made enslavement a central cause of the Civil War as America faced its “Negro problem” after World War II. In the 1970s, America returned again to enslavement, this time in the light of the civil rights movement. Today, the rediscovery can be seen in the new works on the history of capitalism, the “1619 Project” (Hannah-Jones 2021), and the substantively unrelated, but politically comingled, debate over “critical race theory” and the teaching of American history in schools.

Economists have played a key role in some of these rediscoveries, although we typically do not wear this as a point of disciplinary or professional pride. The traditional approaches taken by economists to the study of American enslavement have often, strangely enough, lacked a focus on race. The economic research has focused on agricultural production functions and what they implied about the profitability of enslavers or on patterns of 19th century GDP growth in the North and the South.

■ *Trevon D. Logan is Hazel C. Youngberg Trustees Distinguished Professor of Economics, The Ohio State University, Columbus, Ohio. His email address is logan.155@osu.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.81>.

In the underlying analytical models, enslaved Americans often become a category—perhaps designated by a subscript after a variable—not a group of people caught up in the social, legal, and economic relations that make up systemic racism.

A central component to understanding enslavement, however, requires us to understand race. As Roberts (2011) has argued, race is not a social nor biological category—it is a political one. This implies that race is inherently an economic category as well, as politics is principally concerned with the production, regulation, and redistribution of resources. When economists avoid directly interacting with race and racism, they fundamentally avoid the centrality of race in American political economy, which is to say that they avoid, by choice, a central aspect of the subject.

The racial dimensions of slavery took shape before America was America—they were in place when colonial Virginia designated that the “children got by any Englishman upon a negro woman” would have the status of the mother in 1662. At the time of the founding of the United States more than a century later, America was quite familiar with racialized slavery and racial inequality. This mindset instituted race-conscious discriminatory policies for generations after enslavement ended. The economic system of enslavement continued to define the legal rights, social position, and economic opportunities of Black people well after Emancipation. The steps from America’s racial enslavement regime to the racial inequality we see today is of course not a direct line, but it has the continuity of a winding road along which America’s economy, politics, social norms, and policy have created a web of enduring racial difference. The racial nature of slavery, the structural racial inequality that has followed, and the US economic system should be a key area of focus in American economic history.

In this essay, I sketch out what has been gained and lost due to the uncomfortable relationship between economic historians and American racial history. I briefly summarize how economic research in the last 70 years or so has sought to understand slavery as an economic institution. At first, economic historians in the 1950s focused on profitability and economic sustainability of Southern plantation slavery, along with its effects on the growth prospects of the South. This rediscovery period overturned many Confederate-sympathetic interpretations of the antebellum era. Starting in the late 1960s, when rediscovery of enslavement occurred in the shadow of the civil rights movement, economic historians focused on uncovering primary data about the operation of the slave economy: for example, through detailed examination of records from the 1860 census and plantation records. As new and detailed primary data became available, the earlier questions were then reexamined. These topics remain the subject of ongoing controversies, which I will not seek to resolve here.

The latest work, which arguably started in the mid-1970s but has not yet captured the attention of most economists, involves the recovery of Black economic history. This focuses on questions steeped in actual Black experience. An overly heavy reliance on quantitative data, critical for the earliest work, has limited the questions that economists ask about the relationship between America’s racial and economic history; moreover, the implicit claims of the objectivity of

quantitative data do not always hold up when placed in the context of the racial stain of American history.

Race is not a variable in an econometric model or in a dataset, either—it is also a dynamic experiential condition that is both acted out and acted upon. The path forward requires wrestling with the promises and pitfalls that come with acknowledging that. As an example, I use my family’s narrative and experience as sharecroppers and cotton-pickers in the 1950 and 1960s as a lens and show how answers to questions from economic history and economic theory can be reframed by expanding our evidentiary base and methodological approaches. In the process, I highlight some areas of what these “traditional” economic perspectives miss.

Finally, I briefly provide some examples from other fields—particularly work by historians—that have sought to provide texture on some of the key dimensions of slavery and racial inequality that have been understudied by economists.

Three Phases

Modern research on the economics of American enslavement was more than 20 years old when Fogel (1975) provided a timeline of the research as occurring in three phases. The First Phase, starting in the late 1950s, focused on three inter-related topics: “the profitability of an investment in slaves to slaveholders; the economic viability of the slave economy; and the effect of slavery on southern economic growth” (p. 37). The general thrust of this work was an attempt to overturn a narrative commonly promulgated by historians starting in the 1890s, which claimed that enslavement was conspicuous consumption by a small class of wealthy White landowners and ultimately an unprofitable and unsustainable enterprise, and overall a benign institution. However, the initial findings of Conrad and Myer (1958), using prices for enslaved people and the crops they produced, showed a large, positive rate of return on investment in enslavement, which led to an outpouring of research that borrowed heavily from the burgeoning literature in finance. Yasuba (1961), for example, moved away from rates of return and focused instead on capitalized rents, focusing on the present discounted value of enslaved productivity relative to expenses. Bergstrom (1971) went further and noted that the changes in institutional structure of enslavement had created a capitalized rent *at birthright*—an enslaved person was valuable even before they could be productive. Along with an expanded understanding of the role of violence and terror in plantation management by traditional historians, greater attention to frequent resistance to enslavement among the enslaved, and the effect of enslavement on socialization and personality among the enslaved (Stampp 1956; Elkins 1959; Blassingame 1972; Fogel 2003), the new economic narrative that plantation slavery continued to be profitable and sustainable right up to the US Civil War was a major contribution to the historiography of American enslavement (Easterlin 1960).

The Second Phase of economic research on the institution of slavery, Fogel (1975) argued, was “the massive effort to uncover primary data bearing on the

operation of the slave economy” (p. 38). One prominent example was the efforts of William Parker and Robert Gallman (1992), who dug into manuscript schedules of the 1860 US census. As Fogel notes: “By 1968 they had not only retrieved, but put in machine-readable form, information on 5,000 southern farms and 40,000 slaves” (p. 38). As a result, overall population data from the 1860 US census could be matched with agricultural data from the census on acres of improved and unimproved farmland, values of land and farming equipment, livestock and main crops, as well as number of enslavers and the enslaved. Additional data using probate records, plantation accounts, military records, and birth and death records painted a fuller picture of the agricultural United States before 1860. With this data, new issues could be analyzed, including the distribution of wealth in the North and South, the self-sufficiency of plantations, economies of scale, and calculations of relative efficiency between Northern and Southern farms. It became possible to move beyond rates of return, regional income estimates, and capitalized rents, and to investigate the economic functioning of slavery with microeconomic data.

These issues formed the basis for *Time on the Cross* (Fogel and Engerman 1974), which in two volumes presented work on the operation of the enslaved economy. In many ways, this work crystallized and extended the core issues in the second phase of the research agenda. The central conclusions were that the large plantations in the South were primarily self-sufficient, that there were extensive economies of scale on large cotton plantations, and these economies of scale were a source of efficiency for Southern agriculture overall. Fogel and Engerman argued that the key source of the efficiency advantage for Southern agriculture was the “gang labor” system, which involved a high degree of specialization among relatively small groups of enslaved people in the planting, growing, and picking seasons (Haskell 1975; Gutman 1975; David et al. 1976; Fogel and Engerman 1977; David and Temin 1979; Wright 1979; Fogel and Engerman 1980). This allowed for an incredibly fast pace of work under arduous conditions, and a routinized system that could speed up planting and harvesting to grow and bring to market remarkable amounts of agricultural goods.

Overall, Fogel and Engerman (1974) found that an enslaving farm in the South produced more output per normalized unit of input than either a free farm in the South or the North. Further decomposition showed that this advantage, nearly 33 percent more normalized output, was due to differences in the “labor input.” That is, the South’s economic advantage was principally due to enslavement. As they famously noted, the enslaved produced in 35 minutes what a free farmer in the North took an hour to produce.

It would be difficult to overstate the magnitude of the controversy triggered by Fogel and Engerman’s (1974) *Time on the Cross*. A group of prominent economic historians went through the two volumes and checked every quotation, every piece of data, and every calculation (David et al. 1976; Haskell 1975). The analytic methods used by Fogel and Engerman were also called into question. For example, economic historians engaged in heated disputes over the nature of the efficiency calculation (which used a standard Cobb-Douglas production function), the use of 1860 as

the year for prices and yields, and the disaggregation of farms by type. Additional concerns were raised about the normalization process itself—how can economists compare productivity of Northern farms, which could not grow the most profitable cash crop (cotton), to those that could? There were disputes over whether Southern agriculture actually displayed economies of scale and about whether the gang labor production process played a central role in production, or whether it existed in any widespread way. There was also the curiously absent discussion of the moral issues of American enslavement, and the issue of race more generally.

In a tour-de-force article in the *American Economic Review*, Fogel and Engerman (1977) pushed back against the *Time on the Cross* criticism, primarily defending their productivity calculation. When their critics levied a series of new critiques to their work, most notably David and Temin (1979) and Wright (1979), Fogel and Engerman (1980) again replied in a particularly forceful display. They showed that not only did large Southern plantations produce a tremendous amount of cotton, they produced more than 95 percent of the *non-cotton* output that a Northern farm did as well—implying that Southern farms were indeed on a higher production possibility frontier than Northern farms, which they termed the technical advantage of Southern agriculture. Moreover, that advantage was only seen on enslaving farms.

I will not seek to explicate or resolve these ongoing controversies here. By the end of the 1980s, Fogel had published a four-volume follow-up work, *Without Consent or Contract*, that provides a much more qualified and nuanced view of many of the topics—and also attracted far less controversy. For various overviews of the historical disputes, useful starting points would include Fogel (2003), Wright (2013), and Sutch (2018). At a minimum, it seems fair to say that the American slavery was not doomed by economic forces before the Civil War (Calomiris and Pritchett 2016; Steckel 1986a, 1986b, 2007). More controversially, Baptist (2014) and other “new” historians of American capitalism have argued that the technological advantage of slave-based Southern plantation agriculture and its attendant innovations played a central role in overall American economic ascendancy during the 19th century.

Here, I want to focus on a different aspect of these disputes over the technological status of enslaved-based Southern agriculture: namely, the way in which the assumptions needed for quantitative economic modelling can blur into unwarranted and unsupported assumptions about underlying cultural and social patterns. Fogel and Engerman (1974), as part of their effort to explain the high levels of productivity they found in Southern plantation agriculture, argued for a cultural model of enslavement that would be consistent with its high level of productivity. The obvious attributes of slavery—forced extraction of effort, slow work, resistance, and violence—seemed incompatible with high levels of output produced. Therefore, the Fogel and Engerman model was built on highly productive enslaved workers who responded to “positive” labor incentives, were adequately fed, clothed, and housed, were sexually prudish, who were infrequently punished, where family breakup was rare, and who, overall, exhibited a “Protestant work ethic” that led to the high yields seen in the productivity calculations.

It is difficult to imagine these arguments being made today, and this social model of enslavement has not endured. It was the subject of intense debate and embarrassing refutations of the evidence used to support many of these claims. John Hope Franklin (1980, p. 521), the doyen of African American history, noted that Fogel and Engerman “make claims for the mitigating factors in slavery than they are unable to prove.” Gutman (1975, p. 2), a social historian of enslavement and the Black family, was more forceful:

Whatever its importance as economic history, detailed examination of [*Time on the Cross*] its major arguments and the evidence supporting them shows convincingly that it is poor social history, that its analysis of the beliefs and behavior of “ordinary” enslaved Afro-Americans is entirely misleading, that it uses a thoroughly inadequate “model” of slave socialization, and that it contains frequent and important errors of all kinds in its use of quantitative (not to mention “literary”) evidence essential to its major arguments.

The cultural and social model of enslavement presented in *Time on the Cross* remains a cautionary tale for the limits of quantitative history in general and economic history in particular. What Gutman (1975) notes as a failure of the cultural model of *Time on the Cross* is actually a key feature in deductive economic history reasoning to this day. Economic historians continue to infer meaning from rates and human action through the lens of rational models of behavior, consumption, and production. For example, consider the basic idea that high rates of purchase imply high demand for a good. Whether this conclusion is true or not—that is, whether economists can draw inferences from rates or model estimates to imply something about behavior—depends critically on the context in which that data was drawn and whether that model has any applicability to the people it is hoping to describe. Historians and sociologists might not be able to “count” as rigorously as economists, but they are also highly unlikely to *discount* the context (and in this case, the issue of race) where these decisions are made (Johnson-Hanks 2007). Gutman’s central question of Fogel and Engerman (and of economic historians of enslavement more generally), remains unanswered to this day: “What did the enslaved know and learn, and who did they learn it from?”

Answering this question has plagued the third phase of the research agenda, which Fogel (1975) termed “the recovery of black history.” A Black economic history requires dealing with, principally, Black people, Black culture, Black institutions, Black consciousness, and the appropriate historical context. Given the lack of a nuanced understanding of race, the lack of models centering racial identity, and norms that have treated features of racist systems as “irrational,” there has been little advancement in this area. Instead, economic research on race and racism has tended to concentrate on labor market and housing discrimination, investments in human capital, migration, and policy evaluation. Economic historians have not and do not principally deal with questions that arise from the narrative record, nor do they interface with Black culture, Black institutions, Black historiography,

or racialization in generative ways. There is very little work that begins with Black people's understanding of the world in which they live, as opposed to models that lack an explicit racial element and instead assume that Black and non-Black behavior are the same and perhaps people of different races face different budget constraints. The disconnect between Black economic history and Black narrative history has progressed to the point where Black economic history deals largely with issues from urban and labor economics, and Black narrative history considers larger questions of race, racial formations, culture, intellectual history, and the integration and exclusion of Black Americans from areas of economic and social life.

Moving forward into this Third Phase would require a renewed focus on questions steeped in the Black experience. It is not clear whether the economic history profession, which remains overwhelmingly White and currently preoccupied with "rigorous" tests of models of economic behavior, causal estimates, and/or census linking, has the taste or skill for such analysis. Perhaps self-consciously, much of the work in recent decades on enslavement has turned away from pontificating about cultural and social models—which may be a good thing—to ignoring social and cultural models altogether, which in turn leaves a large gap in our understanding of the economics of enslavement and of race more generally.

As one example of trying to fill this gap, consider the implications of Cook, Logan, and Parman's (2014) discovery of historical Black names. While modern applied economists had developed a robust resume audit literature using given names to infer race and detect racial discrimination in labor markets, the enterprise rested on the faulty presumption that Black names were a product of the civil rights movement. As it turns out, the history of Black names was an inaccurate one. Black names have changed over time, going from Moses and Purlie to Tyrone and Jamal, but they have existed since the first decades after Emancipation. Moreover, Cook, Logan, and Parman (in press) find that enslavement itself may have played a role in the development of historical Black names. While the names were largely race-neutral in the late eighteenth century, the names became Black over the nineteenth century due to both increasing use among Black Americans and decreasing use among White Americans. That is, these names grew in their Blackness as American enslavement reached its zenith.

At the same time, the comparative advantages of economic history are: 1) the ability to derive identification from institutional, historical, narrative, and other qualitative sources—in addition to econometric techniques; 2) to ask and answer questions which are inherently historical as opposed to only economic; and 3) to tell compelling stories that move beyond models to describe the experiences of people.

Every time the history of American enslavement is rediscovered, economic historians find new ways to make enslavement about concepts (coercive labor) as opposed to people (the enslaved). Placing people at the center requires a more humanistic economic history and would move us forward in thinking about race as well. I believe that a nuanced approach to race that emphasizes race as an experience in American history as opposed to a category or a classification goes to the

heart of the epistemology of Black (and American) economic history (Logan 2015). Borrowing liberally from that work, I argue there for the advantages of new approaches that explore and theorize from experiences, which is inherently qualitative work. A new and more relevant Black economic history must include an appreciation of the oral tradition, folklore, and social norms of African American people—the fullness of the Black experience (Levine 1977). At the same time, it must analyze how the economic experiences of Black people have shaped and altered that full experience. At a minimum, the failure to incorporate such experiences implicitly assumes that they are unimportant in economics, which leads us down a path where there is an economics of race about people who have no racial identity that is shaped by the economy.

Lessons to be Learned

What, exactly, can we learn from these alternative approaches? To provide several examples, I return to the terrain of Logan (2015) and consider the issues related to productivity in agriculture. The first issue is gender differences in productivity. Since at least Boserup (1970), differences in men and women's capacity to perform agricultural labor has been linked to a number of patterns of development and technological change. Recent scholarship in economics has sought to exploit gender differences in productivity and their links to economic development around the globe (Alesina, Giuliano, and Nunn 2013). In general, the hypothesis is that geographic differences in crops whose cultivation is more or less gender-equitable can have a profound influence on the scope for technological change, industrialization, household formation, and economic growth (Goldin and Sokoloff 1984). The basic idea is that significant gender differences are a source of comparative advantage for one gender over another to move into light manufacturing and other non-agricultural tasks.

In American economic history, the academic debate has centered around differences in agricultural productivity in the pre-Civil War era (Whartenby 1977; Conrad and Meyer 1958; Parker 1979; Fogel and Engerman 1974, 1980; Fogel 2003; Field 1988; Wright 1978, 2006). Economic historians have estimated gender differences in manual cotton picking, using data from the Second Phase of the research agenda. Unlike other agricultural tasks, cotton picking poses few inherent reasons to suspect significant gender differences: that is, the actual picking of cotton does not require great physical strength, height, weight, or other physical traits that would favor one sex over the other. Also, the record-keeping on plantations in the antebellum era provide a wealth of data to estimate gender differences in manual cotton picking with some precision (Parker 1979; Olmstead and Rhode 2010).

Cotton picking did not end after Emancipation, but records of individual cotton-picking are assumed to be sparse. That may not be the case. There are literally tens of thousands of people, predominantly Black people, who are still alive and who were active in the manual production of cotton well into the 20th century, nearly a century after the end of chattel slavery in the United States. Expanding

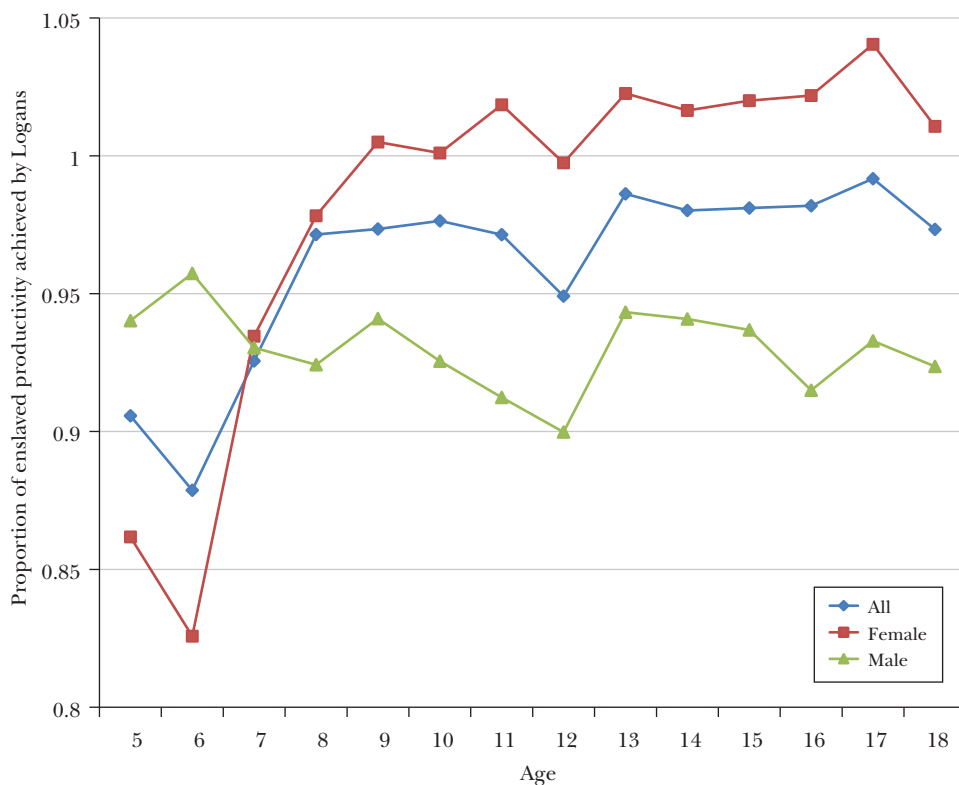
the sources and data and using qualitative techniques can resolve several issues, and point to unexplored features of manual cotton picking that researchers have neglected. I use individual-level data on daily rates of cotton picking from a single, large family observed over a number of years in mid-20th-century sharecropping, inherently controlling for family and other contextual factors that could bias other estimates if they were concatenated together. This investigation is a detailed one based on the experience of one family—my own.

The records recorded here are necessarily specific, but this is common in historical analysis of agricultural productivity: for example, Karakacili (2004) uses manorial roll accounts to estimate agricultural productivity before the Black Death. The cotton picking books I use are the daily records of the pounds of cotton picked, per person. The cotton picking data is a quasi-panel dataset, containing observations on individual cotton pickers at several points in time. I estimate average productivity as a function of age using individual fixed effects. The methodology is similar to Olmstead and Rhode (2010), with the allowance that the dates of picking are not always observed.

At every age, the Logan children picked as much cotton per day as their enslaved forbears. This is seen clearly in Figure 1, which shows the percentage of enslaved productivity achieved by the Logan children at each age overall and by gender. The figure shows that male children in the Logan family were always between 90 and 95 percent as productive as enslaved male children. Females, however, were relatively less productive than enslaved females at early ages, but become more productive than enslaved females by age 10. By the middle teen years, the Logan women were consistently more productive than enslaved women, and differences by gender in productivity were small as well. The difference between early and late teen years is less for Logan children than those who were enslaved, which suggest that Logan children entered into prime-picking productivity at slightly earlier ages than enslaved children.

Using qualitative techniques, I answer an additional question: how was the cotton picked? As noted earlier, there is significant debate in economic history about the use of “gang labor” in cotton agriculture. While Fogel and Engerman (1974) argued that the gang labor system was instrumental in the productivity of enslaved labor, other scholars have questioned whether the system was widely used and the extent to which a gang labor system could explain agricultural productivity on plantations. However, the debate over the use of the gang labor system in cotton picking (the “technological innovation in Southern agriculture”) has not exploited the fact that cotton was picked after Emancipation, and some evidence on the techniques involved in cotton picking after Emancipation could inform the debates on antebellum cotton picking. While the Logans did not pick cotton under chattel bondage, the actual process of manual cotton picking underwent relatively little technological innovation: that is, while there were significant biological innovations in cotton agriculture, the actual manual harvesting process did not change significantly. The qualitative interviews show that the cotton picking process used by the Logans was organized and ordered, but independent (Logan 2015).

Figure 1

Percentage of Cotton Picked per Day by Slaves Achieved by the Logan Family

Source: Logan (2015).

Everybody got two rows—you picked at your right and at your left. You had to zig zag it down. When you was done you went on and got another two rows. When we was real little you picked behind Dad ‘cause you would get the cotton close to the ground, but that didn’t last long. By the time we was in school [age 6 or 7] you had your own two rows and you was out there picking like grown folks.

You had your own stuff to pick. You had your own sack and your own cotton to pick. You didn’t get on nobody’s row and they didn’t get on yours. You had to get to picking. It wasn’t like we was gonna run out of cotton to pick. You got up there and you went at it and that was that. And I tell you, we picked it clean. You had to pick it clean and you had to pick it quick.

In general, the qualitative evidence is not consistent with a gang labor style of organization. One could argue that gang labor was not necessary to achieve high

levels of productivity in manual cotton picking. The organization of the work is only one element of the process, however. The picking of cotton itself took a physical toll on the young children, low levels of productivity were discouraged through punishment, and the labor effort required was high.

The Logan children were involved in all parts of the cotton production process. They planted, chopped, fertilized, and applied pesticides. The interviews stressed that the entire production process was extremely taxing physical labor. Some of the children described chopping—the process of clearing weeds from the cotton plant with a chopping hoe until the plant bloomed—as the most taxing work both due to the process and the environment.

You got a chopping hoe and you got a row. You had to chop every weed around that cotton—you can't do that with two rows and you got to chop it up. But chopping is hard, cause it's hot out. It is hot out. All you would see out there in June and July is Black folks chopping cotton. And when you finished you had to go right back, you just chopped to chop again and again till you see them blooms. That would be about the end of July, a few times earlier. But until you saw them blooms you was out in those fields and you was chopping sun up to sun down. That was about the worst part to me. Then they came out with "freemud" which was some type of poison that killed the weeds. But I was almost done with school when that came out. You didn't have to chop as much then. I remember that I was sixteen when that came out [1965] because I didn't have to work so hard. That was later on.

Due to cotton's long growing season (cotton planting in March or April yields an August to September harvest), chopping comprised a significant portion of the physical labor involved in cotton production.

I couldn't stand to chop no cotton. Those weeds would come up around there and you'd have to whack at it and whack at it. And that sun was out there and you'd be almost dead out there chopping and chopping and chopping. And we had a good little bit of land so we had to get it good and chopped because if them weeds take hold you are messed up. So we had to get it all chopped because you had to go back. I hated it 'cause you was out there all the damn day bent over chopping in the sun. Now, Mississippi in July ain't nothing to play with, and you out there chopping and carrying on and you bent over and chopping at the ground.

This discussion reveals a few facts that are important to the debates about cotton production described above. First, the actual picking of cotton is not seen as the most demanding task. Rather, the chopping of cotton through the growing season was universally noted to be the most arduous work. The focus on the output, measured in pounds of cotton, may have placed undue attention to the activity most temporally related to it—picking.

A second and broader issue is that a standard labor economics approach would not account for this labor, which is a significant portion of their work history and productivity. The productivity of the Logan children in the 1950s and 1960s and many others in manual cotton-picking during this era is far beyond the regular aspects of child labor in a modern industrial economy, nor is it “employment” in a family business, and yet in standard data, it was neither recorded nor noted as part of their economic lives. An individual wage series will not capture this experience and Black/White wage ratios neglect this experience as well. A survey question that asks for employment history would not capture this as employment at all. So what was it? Our standard measures of labor market experience work to make this labor, an inherently racial experience, invisible. This only acts to reinforce the idea that this work is insignificant since we, as economists, do not account for it nor inquire about it when analyzing trends in economic racial inequality.

At an even deeper level, racial inequality may also factor into primitives that economists have taken as given: for example, preferences about savings and rate of time preferences can be the product of racialized experiences. Consider the experience of a 12 year-old in the family in 1961, when a sharecropping family needed loan of \$10 from the landowner. In a sharecropping contract, there is a set amount of money allotted for sustenance, but in a family with 16 children it sometimes was not enough. This loan of \$10 would be paid back through the proceeds of the sale of cotton or could be paid back sooner if members of the household did additional work. This work could include picking peanuts or strawberries, selling wood they chopped, digging a ditch or well, or other manual tasks. Simple financial transactions, however, carried an additional weight of racialized power in the economic system.

I will never forget we needed ten dollars to get some flour and sugar, so Ma sent me and James to Sides [landowner]. [That would purchase two 10-pound sacks of flour (\$6), a 10-pound sack of sugar (\$3), and a canister of coffee (\$1).] He told us “I can’t let you have nothing like that.” And he had a whole hand full of silver dollars in his hand while he did it. He had it right there in his hand. Wasn’t nothing to him. Now, he was sho’ ‘nuff going to get his money back, but he wasn’t going to give us none. And what was we gonna do? We needed to eat. So I made it up in my mind that I ever made me some money I wasn’t ever going to go to some White man for nothing. Even when they knew they was going to get they money they was going to do you bad. People did not have to let you have they money, even if you was good for it and they knew you was good for it. Even if they knew they could get it back. That’s just how it was.

The lack of credit use and over-saving in adulthood by this member of the family is partially due to this racialized experience. At a time when “racial mistrust” has been applied to a number of settings from health care to education to financial

institutions, the context under which the “mistrust” developed should be made clear. Indeed, our language reveals an intent. Mistrust, by definition, assumes an irrationality as the lack of trust is based on intuition. Distrust, on the other hand, is a lack of trust based on experience. Calling these patterns “mistrust” obviates the need to uncover the racial experiences which form the judgments being made. In doing so, we miss the opportunity to interrogate the racial experiences which lead to this behavior.

Another aspect of this particular vignette merits further examination. Even with repayment of such a pitiful amount contractually guaranteed, the racial hierarchy is enough to be denied the smallest amount of credit for no rational reason. This was about more than money (although it was money that they needed); it was about a world where power was expressed through the denial of needs. What does this imply for economics? Our models of taste-based discrimination typically highlight the *disutility* that Whites have when interacting with stigmatized groups, but fail to capture the *utility* that racists gain from actively mistreating those in stigmatized groups. For the racist, racism is consumption. Confining economic concepts to discrimination and avoidance limits the ability of social scientists to describe behavior where racists sought out interracial interaction as a means to exert power, authority, violence, and control.

This insight leads us back to the economics of enslavement. Fogel and Engerman (1974) argued that whippings were rare because positive incentives were used, while more recent work in the history of capitalism stresses the use of whippings as a means of increasing cotton production (Baptist 2014). Both perspectives tacitly concede that whipping is a part of the *production process*, leaving aside the fact that imposing terror may derive utility for the terrorists. Racists may avoid interracial interactions in selected spheres of life, but in other situations racists have desired and draw satisfaction from subjugating Black people, which itself is a social ritual perfected in racial enslavement. The racial order established in enslavement is not only about means of economic production, but also the social reproduction of Black subjugation. Setting aside whatever relationship enslavement has to economic production, the utility derived from racial subjugation deserves a fully articulated economic analysis. For example, it has been argued that racial terror, subjugation, and domination come to form important parts of Black identity that have their roots in enslavement and that form day-to-day practices designed to reinforce a racial hierarchy (Hartman 1997), further extending the theoretical reach of racial enslavement into additional spaces.

The relationship between racial identity and economic identity are deeply related in ways that are immediately obvious in qualitative data but are obscured in much of the current work on race in economic history. Attention to issues of productivity, wages, discrimination, and profit can obscure more fundamental issues of how enslavement formed an economic culture. Race as an experience means that it is a process that is not easily described by a fixed variable in a dataset. Limiting ourselves to the quantitative record gives us partial answers to the questions we ask about racial economic inequality and the endurance of those inequalities over time.

Moving toward the Third Phase

New work from American historians on enslavement awaits the economic history profession. Addressing new questions from the narrative record will go a long way in reinvigorating the Third Phase of the agenda. While economic historians have been active in recent debates in the “history of capitalism” perspective on American enslavement (Hilt 2017), they have confined themselves largely to the topics involved in the first two phases. For example, debates have centered on the growth of cotton yields being more due to increased terror or crop innovation, or the degree to which Southern planters had outsized political influence in international trade. While these are worthwhile, considering the economic implications from innovative scholarship could help move us in more generative directions, and place an emphasis on the people most directly impacted by the peculiar institution.

As one example, consider gender and enslavement. Under the legal doctrine of “coverture,” parts of which continued to prevail into the 20th century, an unmarried woman was allowed to own property and to sign contracts under her own name, but upon marriage, her husband would subsume all such rights. However, despite legal restrictions on married women’s property ownership, enslavement was one area where White women asserted their economic independence. In *They Were Her Property*, Stephanie E. Jones-Rogers (2019) uses a combination of data and narrative to bring a wide range of new questions to the table concerning gender and enslavement. Analysis of records from New Orleans, the largest slave market at the time, reveals that White women were a significant number of the buyers and sellers of enslaved people. Moreover, judges frequently found enslaved people held for “sole and separate use” belonged outside the rules of coverture. Thus, some American White women won economic independence via the ownership of Black people. At a minimum, enslavement as a means around coverture adds complexity to the concept of American slavery as a patriarchal institution.

This insight also raises additional questions: Were creditors regularly duped by husbands, who would later claim that enslaved property was their wife’s possession? In the longer term, how did the perspective of the enslaved on having women as their enslavers influence the ways that Black people interacted with White women before and after Emancipation? How did property rights in people enable a White feminist capitalism among Southern women of the 19th century, which may continue to have racial implications for the women’s movement as it emerged in the later 19th and early 20th century?

The market price of the enslaved changed during their lifetimes, because of age but also as the institution of slavery itself changed. Diana Ramey Berry’s (2017) *The Price for Their Pound of Flesh: The Value of the Enslaved, from Womb to Grave, in the Building of a Nation* details the ways that the value of enslaved people changed over their lives and as enslavement itself changed. Prices in the market have importance for non-market behavior, but this requires a new method of bringing the perspective of the enslaved into the research agenda. How did the enslaved know of their changing value? How did knowledge of value alter relationships on plantations,

family relations, and enslaved fertility and mortality? In addition, given that slaveholders owned the human capital of the enslaved, how did the market value of investments in enslaved skills factor into management decisions? How did the behavior of enslavers in Virginia, a state that grew little cotton, change as cotton grew to be a dominant crop and prices for the enslaved rose? Prices are one part of the story, but a qualitative exploration can answer a broader set of questions related to social and economic relations.

The slave-holding South included not just plantation slavery, but also poor Whites who owned neither slaves nor land, but whose lives were deeply affected by their proximity to slavery. In *Masterless Men: Poor Whites and Slavery in the Antebellum South*, Keri Leigh Merritt (2017) shows that the ascension of the cotton South in the mid-19th century was contemporaneous with the descent of poor Whites into landless poverty. The rise of large plantations and the internationalization and profit in cotton was an economic, social, and political change in American enslavement. For example, Merritt considers the Panic of 1837 and depression that followed. As part of President Andrew Jackson's efforts to shut down the Bank of the United States, he withdrew \$10 million from the bank and deposited it with state-level banks and other financial institutions, which triggered a wave of inflationary lending. Jackson followed up with the Specie Circular executive order in 1836, which required that all sales of government land be conducted in gold or silver. Together with various international economic pressures, major New York banks faced a run where they were unable to provide specie when requested, and the ensuing panic shut down hundreds of banks across the country, with the effects felt disproportionately in the South. Merritt discusses how these events forced poor Whites from their small farms and further increased the cost of land ownership, dramatically increasing the number of poor and landless Whites in the South.

Consistent with Merritt's (2017) analysis, Rhode (2021) analyzes historical census returns and documents an increasing concentration of slaveholding in the South at this time. Earlier histories have assumed a static set of race relations in the South, but Merritt argues that the tools that enforced racial discrimination after the Civil War—disenfranchisement, low investment in public goods, high levels of inequality, disproportionate rates of incarceration—were often deployed against poor Whites in the South before the Civil War, and became racialized tools after the Civil War. From this perspective, the Southern Homestead Act of 1866, which offered government land for sale in Alabama, Arkansas, Florida, Louisiana, and Mississippi, can be viewed not primarily as opportunity for freed slaves (who often found themselves excluded from such purchases), but rather as a reclamation project for poor Whites removed from their land decades before. We know that poor Southern Whites did not gravitate to the antislavery Republican party after the Civil War. More broadly, how much of what economic historians consider preferences are strategic responses to changes in economic institutions? What did the formerly enslaved think and learn about the nature of changing race and class relations, both before and after Emancipation?

Large slave plantations often kept detailed accounting records as part of their management practices. Counting with precision did not end at Emancipation, but the scope in agriculture declined while the scale in industry expanded. Caitlin Rosenthal's (2018) *Accounting for Slavery* argues convincingly that management practices at the micro level in the United States owe a great deal to the basics of plantation management. This raises a series of questions. How did the combination of numeracy and illiteracy interact with economic position for Black Americans in a regime where productivity was individually measured? How did industrial enterprises in the post-Emancipation South capitalize on the familiarity of the formerly enslaved with detailed record-keeping? How were management practices seen as racialized economic interactions? How does resistance to obsessive and aggressive management use of quantitative methods of worker control, from the large industrial enterprises of the late 19th and early 20th century up to the present day, grow out of the use of such tools during slavery?

Moving towards a complete Black economic history will require new methods and new perspectives on race itself that fall outside of traditional economics and even social science more generally. I believe that a qualitative approach is likely to be highly useful. Economists have a tendency to roll their eyes when they encounter "qualitative" as a way of organizing research, but qualitative evidence is *empirical* evidence. It can hold the key to answering questions that quantitative data simply cannot. We should certainly let the quantitative data speak, but the current dominance of census, administrative, business and survey data—and the categories that are used for collecting such data—leaves questions of economic importance unanswerable in the current paradigm. Alongside that data, we can return to the central question: What did Black people know and how did they know it? Progress on the Third Phase requires methods that incorporate a broader range of empirical techniques. When we do so, we move toward the recovery of Black economic history.

■ *Many thanks to Rodney Andrews, Lisa Cook, William Darity, Brad DeLong, Shari Eli, Dania Francis, Richard Hornbeck, Borianna Miloucheva, Chantal Smith, William Spriggs, Richard Steckel, Peter Temin, Jhacova Williams, as well as audiences at the Minnesota Economic Association and Minnesota Center for Diversity in Economics.*

References

- Alesina, Alberto F., Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics* 128 (2): 469–530.
- Baptist, Edward E. 2014. *The Half Has Never Been Told: Slavery and the Making of American Capitalism*. New York: Basic Books.

- Bergstrom, T.** 1971. "On the Existence and Optimality of Competitive Equilibrium for a Slave Economy." *Review of Economic Studies* 38 (1): 23–36.
- Berry, Daina Ramey.** 2017. *The Price for Their Pound of Flesh: The Value of the Enslaved, from Womb to Grave, in the Building of a Nation*. New York: Beacon.
- Blassingame, John W.** 1972. *The Slave Community*. New York: Oxford University Press.
- Boserup, Ester.** 1970. *Woman's Role in Economic Development*. London: George Allen and Unwin.
- Calomiris, Charles W., and Jonathan Pritchett.** 2016. "Betting on Succession: Quantifying Political Events Surrounding Slavery and the Civil War." *American Economic Review* 106 (1): 1–23.
- Conrad, Alfred H., and John R. Meyer.** 1958. "The Economics of Slavery in the Ante Bellum South." *Journal of Political Economy* 66 (2): 95–130.
- Cook, Lisa D., Trevon D. Logan, and John M. Parman.** 2014. "Distinctively Black Names in the American Past." *Explorations in Economic History* 53 (C): 64–82.
- Cook, Lisa D., Trevon D. Logan, and John M. Parman.** Forthcoming. "The Antebellum Roots of Distinctively Black Names." *Historical Methods* <https://doi.org/10.1080/01615440.2021.1893877>.
- David, Paul A., and Peter Temin.** 1979. "Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South: Comment." *American Economic Review* 69 (1): 213–18.
- David, Paul A., Herbert G. Gutman, Richard Sutch, Peter Temin, and Gavin Wright, eds.** 1976. *Reckoning With Slavery: A Critical Study in the Quantitative History of American Negro Slavery*. New York: Oxford University Press.
- Easterlin, Richard A.** 1960. "Interregional Differences in per Capita Income, Population, and Total Income, 1840–1950." In *Trends in the American Economy in the Nineteenth Century: Studies in Income and Wealth*, Vol. 24. Princeton, NJ: Princeton University Press.
- Elkins, Stanley M.** 1959. *Slavery: A Problem in American Institutional and Intellectual Life*. Chicago: University of Chicago Press.
- Fogel, Robert W.** 1975. "Three Phases of Cliometric Research on Slavery and its Aftermath." *American Economic Review* 65 (2): 37–46.
- Fogel, Robert W.** 2003. *The Slavery Debates, 1952–1990: A Retrospective*. Baton Rouge: Louisiana State University Press.
- Fogel, Robert W., and Stanley L. Engerman.** 1974. *Time on the Cross: The Economics of American Negro Slavery*. 2 vols. Boston: Little, Brown and Company.
- Fogel, Robert W., and Stanley L. Engerman.** 1977. "Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South." *American Economic Review* 67 (3): 275–96.
- Fogel, Robert W., and Stanley L. Engerman.** 1980. "Explaining the Relative Efficiency of Slave Agriculture in the Antebellum South: Reply." *American Economic Review* 70 (4): 672–90.
- Franklin, John H., and Alfred A. Moss, Jr.** 1980. *From Slavery to Freedom*. 5th edition. New York: Knopf.
- Goldin, Claudia D., and Kenneth L. Sokoloff.** 1984. "The Relative Productivity Hypothesis of Industrialization: The American Case, 1820–1850." *Quarterly Journal of Economics* 99 (3): 467–87.
- Gutman, Herbert G.** 1975. *Slavery and the Numbers Game: A Critique of Time on the Cross*. Chicago: University of Illinois Press.
- Hannah-Jones, Nikole.** 2021. *The 1619 Project: A New Origin Story*. Edited by The New York Times Magazine, Caitlin Roper, Ilena Silverman, and Jake Silverstein. First edition. New York: One World.
- Hartman, Saidiya.** 1997. *Scenes of Subjection: Terror, Slavery, and Self-Making in Nineteenth-Century America*. New York: Oxford University Press.
- Haskell, T.L.** 1975. "The True and Tragical History of 'Time on the Cross'" *New York Review of Books* 22 (15): 33–39.
- Hilt, Eric.** 2017. "Economic History, Historical Analysis, and the 'New History of Capitalism'." *The Journal of Economic History* 77 (2): 511–36.
- Johnson-Hanks, Jennifer.** 2007. "Natural Intentions: Fertility Decline in the African Demographic and Health Surveys." *American Journal of Sociology* 112 (4): 1008–43.
- Jones-Rogers, Stephanie E.** 2019. *They Were Her Property: White Women as Slave Owners in the American South*. New Haven, CT: Yale University Press.
- Karakacili, Eona.** 2004. "English Agrarian Labor Productivity Rates Before the Black Death: A Case Study." *Journal of Economic History* 64 (1): 24–60.
- Levine, Lawrence W.** 1977. *Black Culture and Black Consciousness: Afro-American Folk Thought from Slavery to Freedom*. New York: Oxford University Press.
- Logan, Trevon D.** 2015. "A Time (Not) Apart: A Lesson in Economic History from Cotton Picking Books." *Review of Black Political Economy* 42 (4): 301–22.

- Merritt, Keri Leigh.** 2017. *Masterless Men: Poor Whites and Slavery in the Antebellum South*. New York: Cambridge University Press.
- Olmstead, Alan L., and Paul W. Rhode.** 2010. "Slave Productivity in Cotton Production by Gender, Age, Season, and Scale." Unpublished.
- Parker, William N.** 1979. "Labor Productivity in Cotton Farming: The History of a Research." *Agricultural History* 53 (1): 228–44.
- Parker, William N., and Robert E. Gallman.** 1992. "Southern Farms Study, 1860." Inter-university Consortium for Political and Social Research [distributor], 1992–02–16. <https://doi.org/10.3886/ICPSR07419.v1>.
- Ransom, Roger L., and Richard Sutch.** 1977. *One Kind of Freedom: The Economic Consequences of Emancipation*. New York: Cambridge University Press.
- Rhode, Paul W.** 2021. "The Centrality of Slavery in the American South: New Facts from the Census Complete Count." Unpublished.
- Roberts, Dorothy E.** 2011. *Fatal Invention: How Science, Politics, and Big Business Re-Created Race in the Twenty-First Century*. New York: New Press.
- Rosenthal, Caitlin.** 2018. *Accounting for Slavery: Masters and Management*. Cambridge, MA: Harvard University Press.
- Stampp, Kenneth M.** 1956. *The Peculiar Institution: Slavery in the Ante-Bellum*. New York: Vintage Books.
- Steckel, Richard H.** 1986a. "Weights and Infant Mortality Among American Slaves." *Explorations in Economic History* 23 (2): 173–98.
- Steckel, Richard H.** 1986b. "A Peculiar Population: The Nutrition, Health, and Mortality of American Slaves from Childhood to Maturity." *Journal of Economic History* 46 (3): 721–41.
- Steckel, Richard H.** 2007. "A Pernicious Side of Capitalism: The Care and Feeding of Slave Children." Unpublished.
- Whartenby, Franklee G.** 1977. *Land and Labor Productivity in United States Cotton Production, 1800–1840*. New York: Arno Press.
- Wright, Gavin.** 1978. *The Political Economy of the Cotton South: Households, Markets, and Wealth in the Nineteenth Century*. New York: Norton.
- Wright, Gavin.** 1979. "The Efficiency of Slavery: Another Interpretation." *American Economic Review* 69 (1): 219–26.
- Wright, Gavin.** 2013. *Slavery and American Economic Development*. Baton Rouge: Louisiana State University Press.
- Yasuba, Yasukichi.** 1961. "The Profitability and Viability of Plantation Slavery in the United States" *Economic Studies Quarterly* 12 (1): 60–67.

The Cumulative Costs of Racism and the Bill for Black Reparations

William Darity Jr., A. Kirsten Mullen, and
Marvin Slaughter

A core contribution of economic historians is to bring quantitative estimates to their topics. With respect to American slavery and its ongoing legacy of racial oppression, the concreteness of numerical estimates plays an additional role. The arguments over payment of reparations to Black descendants of American slaves have many dimensions, but one key requirement of any plan for reparative justice is determination of an appropriate dollar amount. Measures can be partitioned into two broad categories: 1) those based upon itemization of the costs to the victims or gains to the perpetrators from specific atrocities; and 2) those based upon estimates of the combined global effects of the atrocities on living descendants of those persons enslaved in the United States.

The first approach requires identification of a comprehensive list of relevant atrocities and injustices, assigning a dollar value to each, and adding them to arrive at an amount for the reparations bill. In the first main section of the paper, we provide examples of the itemization strategy: the wage costs to the enslaved of bondage; the gains to the perpetrators of slavery; the lives taken by White terrorists in post-Civil War massacres and lynching during Reconstruction and legal segregation; the

■ *William Darity Jr. is the Samuel DuBois Cook Distinguished Professor of Public Policy, Sanford School of Public Policy, Duke University, Durham, North Carolina. A. Kirsten Mullen is a folklorist and the founder of Artefactual, an arts-consulting practice, and Carolina Circuit Writers, a literary consortium that brings expressive writers of color to the Carolinas. Marvin Slaughter is the Interim Director of the African American Leadership and Policy Institute, a community-based research and advocacy organization concerned with collective uplift of Black communities. Darity is the corresponding author at william.darity@duke.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.99>.

property destroyed and seized by White terrorists during the massacres; discrimination in the provision of the homebuying supports from the Federal Housing Administration and the Servicemen's Readjustment Bill of 1944 (commonly known as "the G.I. Bill"); and racial discrimination in employment.

The second approach, to which we turn in the next main section of the paper, requires use of quantitative indicators of the global effects of the impact on the victimized community of a trajectory of injustices. We focus on two global indicators: the present value of providing 40 acres of land to freed slaves in 1865 and the current wealth gap between Black and White Americans.

At the outset, we must say we prefer the second approach, specifically using the Black–White wealth gap as the appropriate gauge for the size of the reparations bill. Not only is full enumeration of all the pertinent atrocities a major challenge, marshalling the data needed to assign a numerical value to each is daunting at best and impossible at worst. A more efficient method is to use a summary measure, like the racial wealth gap, to arrive at the total due.

Moreover, although the amount associated with the itemization method likely will be greater than the bill for reparations based upon the current Black–White wealth gap, all atrocities on an itemized ledger are not applicable to living descendants of the enslaved. In fact, it is hard to justify paying those living today for the harms directly inflicted upon their ancestors, particularly enslavement itself. It is legitimate to pay them for ongoing consequences of the harms that limit their lives in the present moment.

In what follows, at several points, we will make compound interest calculations to generate present value estimates of the costs imposed on African Americans by slavery and subsequent acts of oppression. The choice of interest rate for compounding exercises is somewhat arbitrary. In this essay, we opt to use the two interest rates utilized in Craemer et al.'s (2020) extended analysis of the costs of slavery to the enslaved (discussed further in the next section): 3 and 6 percent. A 3 percent nominal rate approximates a risk-free rate over time; a 6 percent rate can be considered an adjustment that takes into account nominal growth in GDP over time. In addition, the 6 percent rate has historical resonance: it was the "interest rate specified in the sales contract between Georgetown University and the Louisiana purchaser of 272 enslaved people in 1838" (Craemer et al. 2020, p. 219).¹

For those interested in a full presentation of the arguments for a policy of paying reparations to the Black descendants of African slaves, two of the authors of the article have published *From Here to Equality: Reparations for Black Americans in the Twenty-First Century* (Darity and Mullen 2020). Indeed, the final chapter of that book presents a detailed plan of redress for the 40 million Black American descendants

¹Some studies in this literature also include estimates using a 9 percent interest rate, which corresponds roughly to the average nominal rate of return on financial assets on the Standard & Poor's 500 from 1880 to the present (Webster 2021). Use of the 6 percent rate compounded over long spans, compared to a 3 percent rate, leads to explosive increases in present value estimates. We do not report the results of using a 9 percent rate, which leads to even more enormous estimates.

of slavery in the United States. However, whether one agrees with a policy of reparations or not, concrete quantitative estimates of the costs of American slavery and racial suppression to those who were enslaved and oppressed should be a central part of our understanding of US history.

Itemizing Some Costs of Slavery and Racial Oppression

Measuring Costs of Slavery

Enslavement of Africans—both those who were brought here by absolute coercion and their descendants born into slavery—was an organic part of the American republic inaugurated with the Declaration of Independence in 1776. In this subsection, we present four different routes for measurement of the costs of American slavery: unpaid wages to the enslaved, financial gains to the slaveholders, the monetary price the enslaved had to pay to purchase their own freedom, and the contribution of American slavery to the nation's economic development. Table 1 provides a summary of the present value estimates calculated at each of the interest rates.

The first procedure considered for calculating the cost of enslavement to those held in bondage is to use the wage rates paid to free laborers at the time. Of course, because the enslaved population was forced to migrate, in the absence of slavery they would not have been present in the United States in the same numbers. Therefore, in the absence of slavery, the wage rate for free labor, presumably, would have been higher (Darity 1990). However, in lieu of an estimated shadow wage in a hypothetical America *without* slavery, use of the actual prevailing wage rate for hired labor during the same interval provides a cautious estimate of the wage costs of slavery to the enslaved.²

Craemer (2015) uses this approach: that is, based on estimates of the hourly wage rate in each of the years 1776–1860 and estimates of the hours worked by enslaved people, he calculates total wages lost over the interval. Thus, the implicit counterfactual here is that slavery would have been abolished as part of the formation of the United States in 1776, so that slaves and their descendants would have become free labor. Due to uncertainty over the exact size of the enslaved population during the course of the Civil War—it was customary for the enslaved to liberate themselves from the plantations and farms and to join the Union army when it passed through

²Craemer et al. (2020, p. 229–30) point out that had slavery been abolished in 1776, perhaps wages of free labor would not have shifted much: “At least in theory, the addition of freed enslaved laborers might not have exerted a net effect on wages, as it would have increased not only the supply of labor but also, in rough proportion, the demand for free labor by former slave owners.” They conclude, “These countervailing forces might have left the prices for labor in the free labor market roughly unaffected justifying the use of free market hourly compensation records to estimate the earning potential of the enslaved, had they been free laborers at the time.” Nevertheless, they view the superior “counterfactual [as] what the wage rate would have been for an African who migrated freely to the American colonies in a world without American slavery. In that world, voluntary African immigrants would have found higher wages due to a greater labor shortage.” With this counterfactual, using the prevailing wage for free labor is, necessarily, a “conservative” strategy.

Table 1

Comparison of the Measures of the Cost of Slavery

<i>Measure</i>	<i>3 percent interest compounded to the present (2021 dollars)</i>	<i>6 percent interest compounded to the present (2021 dollars)</i>
Missing wages 12-hour estimate	\$8.46 trillion	\$860.2 trillion
Missing wages 24-hour estimate	\$14.2 trillion	\$7 quadrillion
Asset value of the enslaved	\$466.5 billion	\$47.5 trillion
Expense of self-purchase (price of freedom)	\$559 billion	\$56.9 trillion
Contribution to American economic development (1839)	\$33.4 billion	\$6.2 trillion
Contribution to American economic development (1859)	\$72.9 billion	\$105 trillion

Source: Various studies discussed in the text.

their area (Du Bois 1935, p. 55–83)—Craemer (2015) stops the calculation in 1860 rather than 1865, which means that his estimate is biased downward.

Under Craemer’s (2015) first scenario, missing wages are computed on the basis of a 12-hour work day, putting to work all enslaved persons five years of age or older. In his initial work of gathering historical data for wages for free labor, drawn primarily from Officer and Williamson (2022), Craemer uses a 3 percent interest rate to estimate the bill for outstanding wages projected up through 2009. Calculating the sum of the present value of each year’s unpaid labor at 3 percent interest rate yields an estimate of \$8.46 trillion by 2021. At 6 percent, the present value exceeds \$860 trillion.

Craemer’s (2015) second scenario is based upon recognition that the entire day of every day of the week for the enslaved, regardless of their age, constituted theft of their time. Here, Craemer estimates the bill for outstanding wages by imputing the free labor market wage to the full 24 hour day, 365 days a year, for all persons enslaved. The 2009 present value Craemer computes, at a 3 percent interest rate, is \$14.2 trillion. Use of a 6 percent interest rate for calculation of the present value yields a stunning total of \$7 *quadrillion*.

An alternative to the wage-based strategy is to estimate benefits to the slaveholders of the ownership of human property. A monetary value of Black subjugation can be calculated by treating prevailing sales prices for enslaved human beings as capturing the discounted present value of the net expected stream of income generated for the slave owner. This strategy was undertaken by Ransom and Sutch (1990), Neal (1990), and Marketti (1990) in putting forward their respective calculations of “the present value of past exploitation.” Their estimates range from a low of \$17.4 billion (Ransom and Sutch 1990) to a high of \$4.7 trillion in 1983 US dollars.

A second approach is to calculate the asset value of the enslaved population to the slaveholders. On December 31, 1860, upon his exit from Congress in anticipation of his state’s secession from the Union, Louisiana Senator Judah P. Benjamin (1860) estimated that enslaved people had a value of \$4 billion to their

owners.³ The present value of that sum is \$466.5 billion at 3 percent interest and \$47.5 trillion at 6 percent interest.

Given the presence of 4 million enslaved persons, Benjamin's estimate corresponds to an average valuation in 1860 of \$1,000. This estimate enables us to use a third approach, calculating the price an individual had to pay to obtain freedom in 1860 dollars.

Prior to the Civil War, the most common way for a slave to obtain freedom was by self-purchase—buying one's way to emancipation (Alchin 2018). While many enslaved people sought to run away, *successful* escape was difficult because of the specialized teams of "slave catchers" and the presence of federal laws requiring the return of enslaved people to their owners.⁴ Owner-based manumission or legal manumission was quite rare. With the arrival of the Civil War, contributing directly to the Union war effort became an effective way of gaining mass emancipation. By the end of the Civil War, 180,000 Black men served in the Union army—about 10 percent of the military personnel.

Enslaved persons did have a variety of ways to accumulate the funds to buy freedom (Alchin 2018). There were sometimes opportunities to work for pay for the owner or for someone else on Sundays, the ostensible day of rest. In some sectors, forestry and mining in particular, there may have been opportunities for piece rate compensation if the enslaved person met more than their quota. Occasionally there were bonuses at harvest time. If an enslaved person had a garden plot, there might be opportunities to sell some of their produce.

The most important avenue for obtaining the funds to buy freedom was the leasing or "hiring out system," where an enslaved person might be permitted to keep a portion of the money. This was more of an option for enslaved persons who were skilled craftsmen. At times, they might even have had the "privilege" of hiring themselves out, depending upon the degree of leverage their unique skills afforded. Of course, because of the imbalance in power, the enslaved person seeking to buy freedom was reliant on the owner to honor the deal.

The price of freedom generally was 20 percent greater than the market price for an enslaved person (Cole 2005). Therefore, in 1860 an enslaved person would have had to meet an average price of \$1,200. The emancipation value of the 4 million persons enslaved in 1860 would have been \$4.8 billion at the going rate for buying freedom. The present value would amount to \$559 billion at 3 percent and \$56.9 trillion at 6 percent.

A final way of measuring the cost of the atrocity of slavery is to estimate the contribution that enslaved labor made to American economic development. Stelzner and Beckert (2021) offer a comprehensive attempt to arrive at such a calculation. Due to

³Benjamin (1860), who served in three different capacities in Confederate cabinet, including Secretary of War, said "our slaves ... directly and indirectly involve a value of four thousand million dollars." Louisiana's official secession from the United States took place on January 26, 1861.

⁴Nordmann, Christopher A. 2009. "Runaway Slaves." Encyclopedia of Alabama. <http://encyclopediaofalabama.org/article/h-2125>.

data limitations, the authors only are able to calculate the contribution of enslaved labor to national output in two specific years: 1839 and 1859. By totaling the effect of slave production at the regional level, Stelzner and Beckert (2021, p. 19, table 2) arrive at measures of the contribution to national output. Slave-grown cotton in the Southern states was so dominant that the produce of enslaved laborers generated about half of the output in the region.

Stelzner and Becker (2021) conclude that enslaved labor in the United States was responsible for 18.7 to 24.3 percent of commodity output growth during the two-decade interval between 1839 and 1859. In both years, slave production contributed to at least 15 percent of total national output. Present value calculations using the lower bounds on Stelzner and Becker's estimates of the contribution to national output by enslaved persons place the numbers at \$33.4 billion (using a 3 percent rate) and \$6.2 trillion (6 percent) for 1839. A similar calculation for 1859 results in estimates of \$72.9 billion (3 percent) and \$105 trillion (6 percent). Again, these estimates account for only *two years* out of the full span of American slavery.⁵

However, by deriving only the *direct* effects of reliance on slave labor in each region, the authors undervalue the full contribution made by enslaved workers. For example, they assign a value of zero to output attributable to slave labor in New England because there were no enslaved workers in those states in 1839 and 1859. However, the textile sector in New England produced 70–75 percent of the nation's cloth between 1831 and 1860 (Temin 1999, p. 50), and its critical raw material was slave-grown cotton from the US South. The Stelzner-Beckert approach does not take into account extra-regional backward nor forward linkages (Hirschman 1958).

All of the methods discussed here share a common shortcoming: by focusing on lost wages, values from a slaveholder's perspective, or output produced by slave labor, none of them consider in a direct way the full costs of conditions of enslavement, like suffering violence, brutality, and loss of freedom.⁶ In that sense, all of these methods result in downward-biased estimates of the costs of slavery to the enslaved.

Measuring Costs of Post–Civil War Massacres, Murders, and Excess Mortality

Between the end of the Civil War and the 1940s, approximately 100 White massacres directed against Black communities took place (for a list, see Table 2). In 1919 alone, for example, upwards of 35 of these assaults took place in locations

⁵Gavin Wright's contribution to this symposium minimizes the impact of slavery on American economic development. However, he does not directly address Stelzner and Beckert's (2021) analysis, nor does he consider Hirschman's (1958) "backward and forward linkages" extending from the slave plantation sector. Wright does say explicitly that his conclusion does not affect the case for Black reparations.

⁶Craemer et al. (2020) offer illustrations of the monetary costs of the denial of freedom based on compensation awarded to Japanese-Americans subjected to mass incarceration during World War II and for the Americans held hostage in Iran for 444 days starting in November 1979. These estimates are extremely different. Under the Civil Liberties Act of 1988, each survivor of the "internment" was given a letter of apology and \$20,000, which amounted to 76 cents per hour of confinement (in 1988 currency). In contrast, the Justice for United States Victims of State Sponsored Terrorism Act of 2015, amended in 2021, specifies \$10,000 per day of captivity for the hostages, which implies an hourly rate of approximately \$416 per hour in 2021 dollars.

Table 2

List of Documented Massacres and Instances of Mob Violence Perpetrated Against Black Individuals, 1863–1950

<i>Year</i>	<i>Locations</i>	<i>Year</i>	<i>Locations</i>	<i>Year</i>	<i>Locations</i>
1863	Detroit, MI		Shreveport, LA		New London, CT
	New York, NY		Slocum, TX		Norfolk, VA
1866	Memphis, TN		Uvalda, GA		Orono, ME
	New Orleans, LA	1912	Forsyth County, GA		Omaha, NE
1868	St. Bernard Parish, LA		Oscarville, GA		Philadelphia, PA
	Opolousas, LA	1917	Chester, PA		Putnam County, GA
	Albany, GA		East St. Louis, IL		Syracuse, NY
1870	Eutaw, AL		Houston, TX		Texarkana, TX
1871	Meridian, MS	1919	Austin, TX		Vicksburg, MS
1873	Colfax, LA		Baltimore, MD		Washington, DC
1874	Coushatta, LA		Bisbee, AZ		Wilmington, DE
	Vicksburg, MS		Blakeley, GA	1920	Hurtsboro, AL
	Barbour County, AL		Bogalusa, LA		Ocoee, FL
1875	Clinton, MS		Cadwell, GA	1921	Denton, TX
1876	Charleston, SC		Charleston, SC		Tulsa, OK
	Ellenton, SC		Chicago, IL	1922	Perry, FL
	Hamburg, SC		Corbin, KY	1923	Catcher, AK
1887	Thibodeaux, LA		Dublin, GA		Rosewood, FL
1888	Danville, VA		Elaine, AK	1926	Benson, AL
	Fort Bend, TX		Ellisville, MS		Susanna, AL
1891	Omaha, NB		Garfield Park, IN	1943	Beaumont, TX
1896	Polk County, AR		Hattiesburg, MS		Detroit, MI
1898	Phoenix, SC		Hubbard, OH		Los Angeles, CA
	Wilmington, NC		Jacksonville, FL		Mobile, AL
1900	New Orleans, LA		Jenkins County, GA	1946	Chicago, IL
1901	Pierce City, MO		Knoxville, TN		Columbia, TN
1903	Evansville, IL		Lexington, NE	1947	Chicago, IL
	Whitesboro, TX		Longview, TX	1948	Vanport, OR
1908	Springfield, IL		Memphis, TN	1949	Peekskill, NY
1910	Little Rock, AK		Millen, GA		Washington, DC
	New York, NY		Montgomery, AL		
	Norfolk, VA		Morgan County, WV		

Sources: Compiled by the authors as noted in Darity and Mullen (2020, p. 216–17, 374–75 n. 37–42); Haynes (1919); Black Past (2022); Rucker and Upton (2006); Voogd (2008).

as far-flung as Omaha, Nebraska; Chicago, Illinois; Elaine, Arkansas; Austin, Texas; Wilmington, Delaware; and Bisbee, Arizona. The bloodshed of that year was so extensive that civil rights activist and composer James Weldon Johnson dubbed the period the Red Summer.⁷ In these riots and massacres, Black people often were murdered or injured and their property routinely was taken by the White attackers.

Due to the general devaluation of Black lives and the political and financial objectives of the White mobs, there was little concern with establishing precise numbers of Blacks killed during the course of the massacres. In the White riot in Atlanta in 1906, for example, estimates of the number of Black deaths range from 10 to 100; the estimates for the 1917 East St. Louis massacre range wildly from 40 to 200 deaths; and to

⁷Sources Select Resources. “Red Summer of 1919.” SOURCES. https://www.sources.com/SSR/Docs/SSRW-Red_Summer_of_1919.htm.

this day the total number of Black deaths at the notorious 1921 massacre in Tulsa, Oklahoma, sometimes estimated at 300, remains unknown (Bentley-Edwards et al. 2018).

Robert Smalls, a Black Civil War hero and South Carolina congressman (from 1875 to 1887), asserted that 53,000 Blacks had been killed by White terrorists from the end of the Civil War through the rest of the nineteenth century. For many years this figure was viewed as exaggerated, but more recent scholarship suggests that Smalls was roughly correct (Egerton 2018). This would have meant about 1,750 murders per annum in the late nineteenth century, or about five Blacks killed by Whites on a daily basis (Darity and Mullen 2020, p. 166). The most recent estimates of Black deaths on the twentieth century lynching trail amount to about 1,500 (Cook, Logan, and Parman 2018; Fox 2020).

A more comprehensive measure of Black deaths is provided by the estimates of excess all-causes Black mortality generated by public health demographers. Benjamins et al. (2021) find between 2016 and 2018, the all-causes Black death rate was 24 percent higher than the White death rate in the nation's 30 largest cities by population size. This implies 74,402 excess Black deaths in those metropolitan areas during that period alone. Their procedure could be extended nationwide for all years in which vital statistics of a reasonable quality are available.⁸

In placing a monetary value on lives lost, a common approach is to compute what is known as the "value of a statistical life." The value of a statistical life seeks to establish the monetary value of incremental reductions in the likelihood of personal safety risk, in particular premature death, based on empirical estimates of how people actually react when confronted with consumer or workplace choices that involve different levels of risk of death. The value of a statistical life has become a widely used instrument in cost-benefit analyses of proposed government regulations by the Environment Protection Agency, the US Department of Transportation, and other agencies. Today the value used by a number of public agencies approaches \$10 million per person (Kniesner and Viscusi 2019).

Using only the Black excess mortality figure for larger metropolitan areas for 2016–2018, multiplied by the current value of a statistical life, yields a total of \$724 billion. In theory, such a calculation could be extended back in time. But there are data limitations blocking sound estimates of total Black mortality, a necessary prelude to calculation of excess Black mortality, especially prior to the 1940s. As Ewbank (1987) writes:

⁸Wrigley-Field (2020) turns the excess mortality computation on its head and asks how many more White deaths would have to occur to reach the *lowest* recorded incidence of deaths among Black Americans. She concludes: "For hypothetical White age-adjusted mortality to equal the lowest recorded Black age-adjusted mortality, about 400,000 to 420,000 excess White deaths are needed [per annum]. For White life expectancy in 2020 [the pandemic year alone] to fall to the level of the best-recorded Black life expectancy would require an estimated 700,000 to 1 million excess White deaths ... The low-end estimate of about 400,000 excess White deaths is about 5.7 times the current confirmed COVID-19 deaths among Whites, representing an 18 percent increase in White mortality from prepandemic levels. For comparison, this estimate implies that, to reach the best-ever prepandemic Black mortality rates, the full US White population would need to experience a level of excess mortality comparable to 90 percent of the official COVID-19 death rate (for all racial groups) in New York City to date."

There are several reasons why it is difficult to document mortality trends among American Blacks. First, registration of deaths did not begin in the United States until late in the nineteenth century and even then it was limited to a few states. It was not until 1933 that the whole of the continental United States was included in the national Death Registration Area (DRA). Second, the states which started registration systems first were northeastern states. The data on the White population in these states has often been used as a rough approximation of the whole country, but the Blacks in these states are not representative of all Blacks. In 1900 the death registration area included 26 percent of the total population, but only 4.4 percent of the Black population. More than 90 percent of the Blacks in the DRA lived in urban areas while only 23 percent of all Blacks were urban. Third, although Blacks and Whites were theoretically covered by the same data collection systems after the Civil War, the data for Blacks were generally less reliable than those for Whites.

Although vital statistics data on mortality by race for the period before World War II are not fully reliable, the existing numbers can be used to develop rough estimates of excess mortality. A reasonable—and conservative—measure of the Black death rate during this period is 1.6 times the White rate (SoRelle 2000). In 1930, the total US Black population was about 12 million people; the corresponding number of excess deaths would have been about 72,000, or a monetary equivalent of \$720 billion in present value. A similar calculation would have to be performed year by year and the estimates summed to arrive at an approximation of the total monetary value of Black excess mortality.

Measuring Costs of Stolen Property

White terrorist attacks on Black communities in the closing decades of the nineteenth century and the opening decades of the twentieth century were often accompanied by destruction and appropriation of Black property by the mobsters. For many of the White riots, it is very difficult to estimate these costs of property damage.

However, the survivors of the 1921 Tulsa massacre did file insurance claims—albeit unsuccessfully—making it possible to construct a rough estimate of property lost due to the destruction of the Greenwood area of that city (Messer, Shriver, and Adams 2018; Canales and Elder 2021). The sum of the insurance claims filed immediately after the massacre amounted to \$1.8 million. Compounding the present at 3 and 6 percent leads to present value estimates of the damages of \$34.6 million and \$611 million respectively. But Tulsa is an unusual case where researchers can access the insurance claims.⁹

By using data collected on site soon after the event by the courageous anti-lynching activist, Ida B. Wells-Barnett (1920), Logan and Darity (2020) were able

⁹Buck Colbert Franklin (1997), the attorney who narrowly survived the massacre and filed insurance claims and lawsuits on behalf of Black Tulsans, was the father of John Hope Franklin, the famed late African American historian.

to calculate a present value estimate of property losses for Black residents of tiny Elaine, Arkansas. In a massacre where an estimated 300 Black people were killed, Logan and Darity (2020) conclude that about \$10 million in property was taken from the town's Black citizens.

Approximately 1,000 Black people were rendered homeless by the 1919 Chicago massacre (Mitchell 2019), data that affords a lower bound estimate of property lost during that White riot. If six persons lived in the average home, then about 165 homes were destroyed. If the average price of a home in Chicago was \$5,000 at the time, lost Black property amounted to about \$830,000 in value. Compounded to the present, at 3 percent the value destroyed would amount to \$16.9 million; at 6 percent, \$316 million.

In cases where Whites appropriated Black-owned properties, arriving at proper estimates of lost wealth requires access to detailed information about deeds of sale and transfer of ownership as well as information about the race of the buyer and seller. This task is demanding. Even with good access to official records of property sales, severe difficulties can arise, including the trustworthiness of the documents. The example of property sales after the massacre that took place in Ocoee, Florida, in 1920 illustrates the scope of the challenge.

The state of Florida was the site of a series of the most violent reactions of white supremacists to efforts of Black citizens to participate in the electoral process and to evidence of Black economic prosperity (Ortiz 2006; Dunn 2013). Perhaps the most brutal of the assaults took place in Ocoee, Florida, on Election Day, November 2, 1920, after Black residents attempted to vote. Like virtually all White massacres, the exact number of Black people killed is not known, but estimates generally run in the vicinity of 50 deaths.

The trigger for the White riot in Ocoee was the particular efforts of Moses (Mose) Norman to cast his ballot. Rebuffed at the polling place, he and other Black voters were told to get documentation of their eligibility to vote from notary public, R.C. Bigelow, who just happened to be out of town on election day, "on a fishing trip" (Zinn Education Project). Bigelow later figured prominently as the individual who notarized a number of the deeds transferring Black properties to Whites, after the massacre, at fire sale prices. Norman was persistent (as described by the Zinn Education Project):

With little other option, most [Blacks seeking to vote] returned to their homes without casting their ballots. Mose Norman would not be so easily deterred. After being turned away that morning in his Ocoee precinct, he rode to Orlando to seek the council of Judge Cheney. The attorney instructed him to write down the names of any African-Americans who were not permitted to vote and also the names of the poll workers who had denied their Constitutional right. Cheney said a lawsuit against the County could be brought to contest this violation.

Norman returned to Ocoee with these instructions, along with a handful of Black citizens again seeking to vote; as you can imagine, things did not go well.

After again being forcibly turned away, he demanded the poll workers names and exclaimed: “We will vote, by God!”

The response by the Klan was a massacre.

July Perry’s home was the site where the White mob believed Moses Norman had sought refuge after recognizing his life was in danger. Perry, “a farmer and labor broker who dared register fellow Blacks to vote” (Hudek 2020), “lived [with his family] on a large estate that included their home and several barns and outbuildings” (Byrne 2014). Norman and his wife, Elisa, were described as “[living] a little more lavishly, enjoying the fruits of the years of hard work farming the 100-acre family orange grove. He drove around the brick-paved streets of downtown in a fancy six-cylinder Columbia convertible with white sidewall tires, silver spokes, and elegant ‘storm curtains’ instead of side windows” (Byrne 2014). Folk history has it that at one point, before the massacre, Norman was offered \$10,000 for his orange groves, an offer he declined (Byrne 2014).

A third wheel among prominent and prosperous Black residents of Ocoee was Valentine Hightower, also the owner of substantial tracts of land, who apparently was less demonstrative about his economic success (Byrne 2014):

Unlike his two friends, Valentine Hightower had a much more low-key personality. He knew well the dangers “negroes” opened up for themselves when they got too “uppity” in the still very white supremacist society of the time. He and his wife Jane lived a modest lifestyle and raised their kids to mainly keep to themselves. They answered politely when spoken to and preferred not to draw too much attention to themselves.

“My father was kind of humble,” Hightower’s son Armstrong recalled many decades later, “He didn’t take no chances on nothing.”

July Perry was the first victim of the mob, captured and lynched after the terrorists were unable to find Moses Norman on Perry’s property. Norman, himself, fled Ocoee and may have left the state of Florida, ultimately settling in New York. Valentine Hightower and his family “relocated” to Florida’s Dade County.

What happened to their property was representative of the aftermath of many white supremacist uprisings. Bluford Sims, a former Confederate officer from Tennessee and one of the major actors in the riot, arranged to be named executor of Perry’s estate and sold off the property at a significant economic gain with the provision that the property never be sold to another Black person (Fussell 2016; Bracy 2019).¹⁰ A search of the records of deeds for Orange County, Florida, where

¹⁰There has been a movement to change the name of Bluford Avenue in Ocoee, named after Sims, to July Perry Avenue.

Ocoee is located, does not provide evidence that Sims purchased the properties from Perry's survivors, who were driven out of the town.¹¹

On the other hand, there is evidence of a formal transfer of property from Norman and Hightower to White ownership. On February 28, 1923, the records indicate that Moses Norman and his wife Eliza, in conjunction with Valentine Hightower and his wife Jane, sold a property to I.C. Bobo for \$10. On the deed of sale, Moses Norman's attestation to his agreement to the sale was verified by a notary public in New York, Joseph L. Pritchard, where Norman, presumably, had taken refuge. The Hightowers sold another property to I.C. Bobo for only \$1 on January 26, 1923. We were unable to find any additional evidence confirming Norman's presence in the state of New York after the massacre. We even remain uncertain about the validity of the signatures of the Black property sellers on the deeds of sale, particularly since it may have been quite dangerous for them to return to Orange County to sign over the properties in person.

Using the census records of Black ("negro" or "colored") residents of Ocoee in 1920 provided in Appendix A of a report by the Office of Program Policy Analysis and Government Accountability (2019) of the state of Florida, it is possible to identify cases where a Black owner is making a "sale" to a White buyer. The property transfers accompanied the permanent exit from Ocoee of many Black survivors of the White terrorist attack. After the massacre, in the climate of coerced sales, our search of the property deeds from that time find that the going price seems to have been \$10, regardless of the size and the structures on the property, as shown in Table 3.

As one example, Jackson and Annie Hamiter appear to have received a more reasonable price of \$200 from notary R.E. Bigelow for a property sold on April 20, 1925, but they also had a succession of four sales between 1922 and 1925 at the \$10 rate. Mrs. Hamiter penned a letter that set the context for the property sales that followed at injuriously low prices (as quoted in McLeod and Dickinson 2001):

On the north side of town all the homes and some of the people were burned, one man was shot and killed and carried to the county seat and hung up as lynched. ... The people on the south of town are being threatened that they must sell out and leave or they will be shot and burned as the others have been. ... It seems to have been a prearranged affair to kill and drive the colored people from their homes as they were getting more prosperous than the White folks.

According to the 1920 US Census, 255 Blacks lived in Ocoee, Florida. The US Census of 1930 counted two remaining Blacks in Ocoee in that year, with the number decreasing to zero between 1940 and 1970.

¹¹The authors want to express their gratitude to Pamela Schwartz at the Orange County Regional History Center and Paul Ortiz, professor of history at the University of Florida, for clarifications about the information contained in the records of deeds.

Table 3

Cross-Referenced List of Black Landowners in Ocoee, FL, Approximate Property Size, and Valuation

<i>Last Name (Alternate Spelling)</i>	<i>First Name (Alternate Spelling)</i>	<i>Property Size (Acres Rounded to Nearest Whole Number)</i>	<i>Purchase Value (Year)</i>	<i>Sale Value (Year)</i>
Anderson	Garfield	79	Federal Land Grant	\$100 (1922)
Battsey(Betsey)	Randolph (Randall D, R.D.)	1	\$25 (1899)	* All holdings sold together for \$10
Battsey(Betsey)	Randolph (Randall D, R.D.)	3	\$60 (1906)	* All holdings sold together for \$10
Battsey(Betsey)	Annie	1	\$25 (1908)	* All holdings sold together for \$10
Blackshear	Martin	1	\$10 (1908)	\$10 (1921)
Blackshear	Martin	6	\$50 (1915)	\$25 (1920)
Blue	Sanborn (Sanders, Sanderson, Sanborri Blue)	3	\$250 (1920)	\$750 (1920)
Thomas	Dennis	5	\$100 (1897)	\$1,000 (1920)
Dighs	Edward	79	\$500 (1919)	No recorded deed of sale
Edwards	John (Joe)	5	\$150 (1914)	\$10 (1921)
Frank(s)	Daniel (D.R.)	20	\$50 (1911)	No recorded deed of sale
Green	Sally (Sallie)	3	\$54 (1916)	\$10 (1923)
Hamriton (Hamiter)	Jackson (Jack, J.H.)	1	\$25 (1905)	\$10 (1924)
Hamriton (Hamiter)	Jackson (Jack, J.H.)	1	\$450 (1920)	\$10 (1925)
Hightower	Valentine (V.T.)	52	unknown**	\$10 (1926)
Hightower	Valentine (V.T.)	2	\$50 (1912)	\$1 (1923)
Johnson (Sr.)	Stephen (Steve)	6	\$50 (1913)	\$10 (1922)
Lynch	Richard	3	\$90 (1898)	\$1 (1921)
Lynch	Richard	1	\$25 (1905)	\$10 (1925)
Mcras (McRhea, McRae)	William (W.M.)	1	\$75 (1919)	\$450 (1920)
Moore	Rocky	X	X	X
Nelson	Stephen	X	X	X
Penger (Penzer)	Perry F (Kerry)	X	X	X
Perry	Julius (July, J.P.)	5	\$100 (1890)	No recorded deed of sale
Perry	Julius (July, J.P.)	1	\$125 (1907)	No recorded deed of sale
Stater	Victoria	X	X	X
Surrency	Jessie (J.C.)	3	\$45 (1915)	\$100 (1920)
Warren (Warron)	Wade	X	X	X

Source: Compiled by the authors from data obtained from Office of the Comptroller, Orange County, Florida (<https://or.occompt.com/recorder/eagleweb/docSearch.jsp>).

Despite having access to records of ownership changes that took place in Ocoee, Florida, after the White riot, it remains difficult to gauge the full loss in wealth to Black residents. It is difficult to determine what the properties were worth at the times of the post-massacre sales. Tax appraisals do not correspond neatly to the sizes of the lots that were purchased for \$1 and \$10 by Whites. However, Table 1 shows that in those instances where there were recorded deeds of sale, sales prices for Black-owned properties were generally far below the earlier purchase price, regardless of the size of the lot.

The partial evidence that is available makes clear that property losses and forced sales were common results of the approximately 100 massacres, but, for most, systematic records are not available. This is another instance in which calculation of the total cost for a particular item on the redress docket may be intractable.

Measuring Redress for Twentieth-Century Discrimination in Homeownership

Today, there is both a racial gap in rates of homeownership—73 percent for Whites versus 45 percent for Blacks—and also a racial gap in the equity values associated with White- and Black-owned homes. Zillow listing prices indicate that a home in a neighborhood with no Black residents has a median value of \$341,000. In contrast, homes in neighborhoods with a majority of Black residents have a median value of \$184,000; these differences persist independent of the structural characteristics of the home (Perry, Rothwell, and Harshberger 2018). The average level of equity Whites hold in their homes is \$216,000; for Blacks, the average level is \$94,000 (Ross 2020).

Several mechanisms denied Blacks access to home ownership in the early part of the twentieth century, which, in turn, denied Blacks access to the benefits of appreciation in home equity gained by Whites throughout the latter half of the twentieth century. Restrictive covenants in property deeds blocked the sale of property to Blacks (and others). The Federal Housing Administration, established in 1934, formed a public-private partnership with local bank mortgage lending and openly practiced “redlining,” which involved not supporting loans to geographic areas with minority population, thus limiting Black homebuying opportunities (Rothstein 2017). Nevertheless, perhaps the single biggest factor leading to the sharp divide in Black and White wealth associated with homeownership was the Servicemen’s Readjustment Act of 1944, commonly known as the G.I. Bill.

The G.I. Bill may have been the most powerful twentieth century vehicle for upward mobility enacted by the US government (Blakemore 2021). The legislation provided support for veterans seeking higher education and vocational training, to purchase farms or start businesses, and unemployment relief. The most important element of the G.I. Bill for wealth-building and entry into the middle class were the homebuying subsidies and low- to zero-interest mortgage loans (Thompson 2019).

But the application of the law was decidedly inequitable, lifting millions of White World War II veterans to the “middle class,” while leaving many Black veterans behind (Katznelson 2005). Government sanctioned residential segregation limited the locations where Black veterans could buy homes—most notorious, perhaps, was their exclusion from multiple Levittown developments (Rothstein 2017).

In addition to restrictions on their ability to use mortgage support from the federal government, Black veterans had limited access to the homeownership benefits of the G.I. Bill in the first place. In 1947, only two of the 3,200 Veterans Administration guaranteed home loans for veterans in 13 Mississippi towns and cities went to Black veterans, in a state where close to 40 percent of the returning veterans were Black; in New York and New Jersey, less than 100 out of 67,000 mortgage loans went to Black veterans (Levinson 2020). By 1956, the final year of the first G.I. Bill, 4.3 million home loans amounting to \$33 billion had been distributed to World War II veterans, overwhelmingly and disproportionately White (Blakemore 2021).¹²

Although the administration of G.I. Bill educational benefits does not appear to have been as inequitable as the provision of the homeownership benefits, discriminatory disparities also were associated with this component of the legislation (Turner and Bound 2003).

The language of the G.I. Bill was race-neutral, but its execution was not. To gain support for the legislation from Southern Democrats, the administration of the G.I. Bill had to be conducted on a decentralized basis by placing control over the program in the hands of state and local authorities. To construct a measure of the magnitude of resources denied Black veterans, it would be necessary to model not just Black veterans who were denied benefits, but also those who did not apply in the expectation of being rejected. Thus, estimates could be generated of the lost income due to the inability of Black veterans to pursue the full range of educational opportunities promised by the law and wealth lost due to exclusion from buying homes in neighborhoods where properties underwent markedly greater appreciation rates than those in predominantly Black neighborhoods.

Measuring the Black–White Gap in Wage Costs

In a contribution to Richard America's edited volume, *The Wealth of Races*, Chachere and Udinsky (1990) provided a method for calculating the costs to Black Americans of labor market discrimination from employment discrimination during the 40-year interval 1929–1969. They worked with the assumption that 40 percent of the racial income differential was attributable to labor market discrimination and reached a calculation of the accumulated benefits to Whites of \$1.6 trillion by the mid-1980s. Projecting their estimate forward in the decades since then, the corresponding value would be \$4.6 trillion (using a 3 percent interest rate) or \$13 trillion (6 percent).¹³

¹²Eden (2021) argues that the difference in nominal monetary outlays per Black and White veteran was not significant, but the use of a cash-equivalent measure of their impact indicates substantial discriminatory effects to the disadvantage of Black veterans.

¹³Swinton (1990) modified the Chachere and Udinsky (1990) estimate by subtracting total government expenditures on social safety net programs, including Social Security, Medicare, Medicaid, and unemployment insurance. Swinton's calculation was a response to the claim often made by anti-reparationists that Blacks already have received restitution in the form of social transfer programs. This claim seems shaky. After all, the majority of funds from social transfer programs since the 1960s did not go exclusively to Black Americans; indeed, for a full 30 years after the onset of the New Deal project, Black Americans,

In what follows, we sketch an alternative strategy for calculating the wage loss to Black Americans from the inherited legacy of past racial discrimination, including not just discrimination by employers but also the broader forms of structural racism that have constrained educational and occupational choices for Blacks. Our estimate consists of two major components: a reduction in wages and a greater likelihood of being unemployed for Black workers. We apply this approach to 2019 data to arrive at a rough estimate of the impact of labor market discrimination on lost income for Black employees and potential employees.

The median hourly wage for White workers was \$21.32 in 2019 (Gould 2020, table 3). The full difference in median Black and White hourly wages was \$5.20: \$21.32 minus \$16.12. Suppose Black workers are assumed to have lost, conservatively, 5 percent of their earnings due to employment discrimination. Then the hourly wage loss for each Black worker was \$1.06.

If we assume the typical full-time worker was paid for 48 weeks, five days a week, for an eight-hour workday, total annual hours would have been 1,920. If there were about 20 million Black labor force participants with a 6.1 percent annual unemployment rate, 18.6 million Black Americans received pay over the course of the year for 35.7 billion hours of work. The difference in the Black and White unemployment rates in 2019 can provide one gauge of the number of Black hours of work lost due to discrimination. The unemployment rate for Whites in 2019 was 3.3 percent. Subtracting 3.3 percent from 6.1 percent leaves a gap of 2.8 percent. Multiplying 20 million by 2.8 percent yields 360,000 Black Americans subjected to *excess unemployment*. Then, multiplying 360,000 persons by 1,920 annual hours leads to a total of 691.2 million hours of work lost due to structural racism.

Combining both the lower wage for employed Blacks and the loss of hours due to the higher unemployment rate for Blacks leads to an estimate of an aggregate loss of \$38.6 billion in earnings in 2019 for Black workers. This would amount to an average loss of \$1,930 per Black labor force participant in 2019 alone.

This estimate is necessarily very rough: for example, it ignores racial differences in discouraged unemployment, typically more of a burden for Black workers. This estimation procedure also does not adjust for racial differences in part-time work when the employee desires longer hours. Still, to the extent that annual calculations of this type can be made for each year from the end of slavery to the present, and then compounding past dollar values to the present, this estimation provides an avenue to arrive at an extensive measure of the cost in wages of structural racism to its victims. However, the data issues here are substantial: for example, 1948 was the first year that the US Bureau of Labor Statistics collected data by race on unemployment.

de facto, were excluded from most benefits of social transfer programs (Katznelson 2005). Nevertheless, Swinton deducted the total cost of social safety net funding between 1929 and 1969 from the Chachere and Undinsky and still arrived at a net \$500 billion gain for Whites from labor market discrimination against Blacks measured in 1985 dollars. This, in turn, would amount to \$1.4 trillion (using a 3 percent rate) or \$4.1 trillion (6 percent) by 2021.

Unified Measures of the Costs of Slavery and Oppression

The itemization approach has the advantage of making the atrocities committed against African Americans concrete. But in many cases, the information simply is not available to construct all of these measures; in others, there will be overlap between different measures. Here we propose a more compact and efficient method for estimation of the invoice for reparations. We discuss two approaches: 1) calculate the present value of a debt that has remained unpaid for more than 155 years, and 2) utilize the Black–White wealth gap as an indicator of the long-term effects of racial injustices.

The Unpaid Debt

Even prior to the end of the Civil War, a series of “experiments” had been conducted with land settlement and independent farming among the formerly enslaved who joined the Union forces as they moved through the south. These experiments had been highly successful in demonstrating the freedmen’s skill and productivity in agriculture; after all, they had been the direct producers of the crops of plantation-based agriculture. The most notable of these was the project undertaken at Port Royal, South Carolina. However, via a variety of machinations, the land was procured by Northern speculators by outbidding the freedmen at auctions with their far greater resources (Franke 2019).

In January 1865, General William T. Sherman issued Special Field Orders No. 15, designating 5.3 million acres stretching from the Sea Islands of South Carolina to northern Florida bordered by the St. Johns River, to be settled by land grants of 40 acres to freed slaves. However, after 40,000 of the formerly enslaved had settled on 400,000 acres allotted under Sherman’s order, President Andrew Johnson ended the project and restored the land to the former slaveholders (Darity and Mullen 2020, p. 156–59).

The Freedmen’s Bureau Acts of 1865 and 1866 also included provisions for land distribution to the formerly enslaved, but these provisions were implemented tepidly at best. The Southern Homestead Act enacted in 1866 explicitly designated the freedmen as beneficiaries of the land patents allotted under the legislation, but the law, which specified 46 million acres of land in Alabama, Arkansas, Florida, Louisiana, and Mississippi be distributed to the freedmen and White Unionists at below market prices, was left to local administration, largely managed by the former Confederates. The law was terminated in 1876. An estimated 4,000 to 5,500 Blacks received federal patents from the Southern Homestead Act before its close, out of four million freedmen (Darity and Mullen 2020, p. 37). Southern historian Walter Fleming (1906), who can reasonably be treated as a hostile commentator on the merits of Black Reconstruction, nevertheless concluded that Black “expectations” of land grants “were justified by the policies of the government and the actions of its agents” (p. 721). But the promise was unfulfilled.

What would have been the value of such land? Consider a back-of-the-envelope estimate predicated on the 40-acre allotments being made, on average, to

four-person households. Since there were 4 million freedmen, the total promised allocation would have been about 40 million acres. An average land price of \$10 per acre in 1865 (Mittal and Power 2000) implies that the projected land distribution would have been worth \$400 million—for perspective, about 10 percent of the value that these human beings were worth to their owners during the period of slavery. The present value of \$400 million in 1865 compounded to the present at 3 percent interest is \$40.2 billion and at 6 percent interest is \$3.5 trillion. Under the conditions of the Homestead Act of 1862, 1.5 million White families ultimately each received 160-acre land grants in the western territories—four times the size of what was promised but not delivered to Black families.

While the average annual rate of real estate appreciation runs in the zone of 3 to 5 percent (Yale 2021), many of the Sea Island properties in the South Carolina that were to have gone to the freedman have experienced appreciation rates in the vicinity of 5 and 6 percent. White land speculators bought many of these properties in the 1860s at less than \$1 per acre.

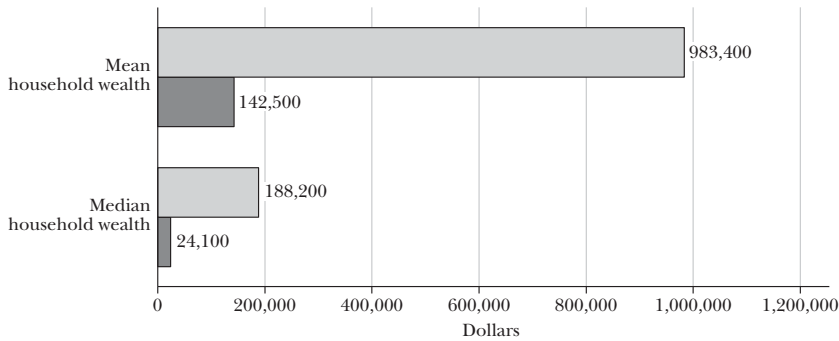
In 2011, former South Carolina governor and Senator Mark Sanford, whose family had been the most recent buyers of a certain plantation, placed it under a conservation easement valued at \$2.5 million (Franke 2019, p. 74–76). Sanford described the property as having been restored to its original status as a “family farm,” eliciting the following stinging comment from Katherine Franke (2019, p. 76): “Mark Sanford’s use of the term ‘family farm’ has such a quaint ring to it, erasing entirely the history of enslavement and failure to make good on the promise of land-based reparations for the people who had been enslaved by the Chisolm family.” Franke describes a number of other properties that were originally designated for freedman, then scooped up by Northern investors, and eventually became estates bought and sold by the rich. Black Americans were deprived of a possible avenue to intergenerational wealth when they were denied the 40-acre land grants.

The Modern Racial Wealth Gap

In discussions of paying reparations, a common question is the extent to which it is appropriate to compensate living African Americans for long-ago harms inflicted upon their ancestors. To assess the contemporary consequences for living African Americans of the atrocities experienced by their ancestors, we believe that the current Black–White disparity in wealth is the best single economic indicator of the *cumulative, intergenerational* impact of White racism over time. Blocked opportunities for Black wealth accumulation are transmitted to current generations via sharply reduced transfers of resources from parents and grandparents to children and grandchildren relative to the transfers made in White families (Darity and Mullen 2020, p. 30–37).

According to the 2019 Survey of Consumer Finances conducted by the Federal Reserve, the Black–White difference in household net worth at \$164,100 at the median and at \$840,900 at the mean, as shown in Figure 1 (Bhutta et al. 2020). The average White household consists of 2.36 persons while the average Black household consists of 2.46 persons; the per capita difference is about \$67,200 at

Figure 1

Two Views of the Black–White Wealth Gap

Source: Bhutta et al. (2020).

the median and about \$358,300 at the mean. There are approximately 40 million African Americans who are descendants of persons enslaved in the United States (or descendants of the freedmen), out of a total of 45 million Black Americans. Therefore, the sum of the Black–White wealth gap across households would be \$2.7 trillion if evaluated at the median, and \$14.3 trillion if evaluated at the mean.

Most discussions of economic inequality tend to focus on the median, because the middle values in a distribution, uncontaminated by outliers, are assumed to be more representative of the typical experience of members of each group. But in this case, we contend that the mean gap is more relevant. The ferocious degree of concentration of wealth in the United States means that 97 percent of the wealth held by all White Americans is possessed by White households with a net worth greater than the median White household. Thus, a focus on the median, rather than the mean, takes virtually all White wealth out of consideration for equality (Darity, Addo, and Smith 2020, p. 499).¹⁴

However, while we view the \$14.3 trillion estimate of what is required to close the Black–White wealth gap at the mean as the single-most appropriate measure of the cumulative and intergenerational impact of White racism on current descendants of slaves, we also view it as a minimum estimate.¹⁵ After all, the Black–White wealth gap results from past experience, but living Black Americans continue to experience

¹⁴ Furthermore, the gap at the mean is not solely due to the presence of a small number of extraordinarily rich White billionaires. One-quarter of White households have a median net worth in excess of \$1 million, a circumstance that is true for only 4 percent of Black households (Darity, Addo, and Smith 2020, p. 498). In addition, members of the White working class have two to three times the wealth than members of the Black professional managerial class (Addo and Darity 2021, p. 182).

¹⁵ For example, Tasseli McKay (forthcoming) constructs specific estimates of the costs of excess incarceration on Black Americans, costs that may be added above and beyond closing the racial wealth gap. However, the financial effects of past over-incarceration also may be embedded, in part, in the Black–White wealth differential itself.

excess mortality, as well as discrimination in employment, credit markets, and the criminal justice system.

■ *The authors wish to acknowledge the support of the W.T. Grant Foundation for a project designing a program of African American reparations and to acknowledge Thomas Craemer and the editors of the Journal of Economic Perspectives for thoughtful suggestions that much improved the article.*

References

- Addo, Fenaba, and William Darity Jr.** 2021. "Disparate Recoveries: Wealth, Race, and the Working Class after the Great Recession." *ANNALS of the American Academy of Political and Social Science* 695 (1): 173–92.
- Albright, Alex, Jeremy A. Cook, James J. Feigenbaum, Laura Kincaide, Jason Long, and Nathan Nunn.** 2021. "After the Burning: The Economic Effects of the 1921 Tulsa Race Massacre." NBER Working Paper 28985.
- Alchin, Linda.** 2018. "Buying Freedom from Slavery." American-Historama.org. <https://www.american-historama.org/1790-1800-new-nation/buying-freedom-from-slavery.htm>.
- Benjamin, Judah Philip.** 1860. "You Never Can Subjugate Us." Speech, US Senate, December 31, 1860. <http://civilwarcauses.org/judah.htm>.
- Benjamins, Maureen R., Abigail Silva, Nazia S. Saiyed, and Fernando G. De Maio.** 2021. "Comparison of All-Cause Mortality Rates and Inequities Between Black and White Populations across the 30 Most Populous US Cities." *JAMA Network Open* 4 (1).
- Bentley-Edwards, Keisha L., Malik Chaka Edwards, Cynthia Neal Spence, William Darity Jr., Darrick Hamilton and Jasson Perez.** 2018. "How Does It Feel to Be a Problem? The Missing Kerner Commission Report." *Russell Sage Foundation Journal of the Social Sciences* 4 (6): 20–40.
- Chachere, Bernadette, and Gerald Udinsky.** 1990. "An Illustrative Estimate of the Present Value of the Benefits from Racial Discrimination 1929–1969." In *The Wealth of Races: The Present Value of Benefits from Past Injustices*, edited by Richard F. America, 163–8. Westport CT: Greenwood Press.
- Bhutta, Neil, Andrew C. Chang, Lisa J. Dettling, and Joanne W. Hsu.** 2020. "Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances." Washington: Board of Governors of the Federal Reserve System.
- Black Past.** 2022. "Racial Violence in the United States Since 1526." <https://www.blackpast.org/special-features/racial-violence-united-states-1660/>.
- Blakemore, Erin.** 2021. "How the GI Bill's Promise Was Denied to a Million Black WWII Veterans." *History*, April 20. <https://www.history.com/news/gi-bill-black-wwii-veterans-benefits>.
- Bracy, Randolph.** 2019. "Ocoee Needs Alternative to Divisive Founders' Day." *Orlando Sentinel*, October 18. <https://www.orlandosentinel.com/opinion/guest-commentary/os-op-randolph-bracy-july-perry-ocoe-20191018-r55k4qzbr5b3xm2ez53ybl6wcu-story.html>.
- Byrne, Jason.** 2014. "Ocoee on Fire: The 1920 Election Day Massacre." *Medium*, November 23. <https://medium.com/florida-history/ocoe-on-fire-the-1920-election-day-massacre-38adbda9666e>.
- Canales, Manuel, and Scott Elder.** 2021. "A Lasting Destruction." *National Geographic*, June, 70–71.
- Cole, Shawn.** 2005. "Capitalism and Freedom: Manumissions and the Slave Market in Louisiana, 1725–1820." *Journal of Economic History* 65 (4): 1008–27.
- Cook, Lisa D., Trevon D. Logan, and John M. Parman.** 2018. "Racial Segregation and Southern Lynching." *Social Science History* 42 (4): 635–75.

- Craemer, Thomas.** 2015. "Estimating Slavery Reparations. Present Value Comparisons of Historical Multigenerational Reparations Policies." *Social Science Quarterly* 96 (2): 639–55.
- Craemer, Thomas, Trevor Smith, Brianna Harrison, Trevon Logan, Wesley Bellamy, and William Darity Jr.** 2020. "Wealth Implications of Slavery and Racial Discrimination for African American Descendants of the Enslaved." *Review of Black Political Economy* 47 (3): 218–54.
- Darity, William, Jr.** 1990. "Forty Acres and a Mule: Placing a Price Tag on Oppression." In *The Wealth of Races: The Present Value of Benefits from Past Injustices*, edited by Richard F. America, 1–14. New York: Greenwood Press.
- Darity, William, Jr.** 2021. "Why Reparations Are Needed to Close the Racial Wealth Gap." *New York Times*, September 24. <https://www.nytimes.com/2021/09/24/business/reparations-wealth-gap.html>.
- Darity, William, Jr., Fenaba R. Addo, and Imari Z. Smith.** 2020. "A Subaltern Middle Class: The Case of the Missing 'Black Bourgeoisie' in America." *Contemporary Economic Policy* 39 (3): 494–502.
- Darity, William, Jr., and A. Kirsten Mullen.** 2020. *From Here to Equality: Reparations for Black Americans in the Twenty-First Century*. Chapel Hill: University of North Carolina Press.
- Du Bois, W.E.B.** 1998. *Black Reconstruction in America, 1860–1880*. New York: Free Press (1935).
- Dunn, Marvin.** 2013. *The Beast in Florida: A History of Anti-Black Violence*. Gainesville: University Press of Florida.
- Eden, Maya.** 2021. "Qualifying Racial Discrimination in the 1944 G.I. Bill." Unpublished.
- Egerton, Douglas.** 2018. "Terrorized African Americans Found Their Champion in Civil War Hero Robert Smalls." *Smithsonian Magazine*, September. <https://www.smithsonianmag.com/history/terrorized-african-americans-champion-civil-war-hero-robert-smalls-180970031/>.
- Ewbank, Douglas.** 1987. "History of Black Mortality and Health before 1940." *Milbank Quarterly* 65 (S): S100–128.
- Fleming, Walter.** 1906. "Forty Acres and a Mule." *North American Review* 182 (594): 721–37.
- Fox, Alex.** 2020. "Nearly 2,000 Black Americans Were Lynched During Reconstruction" *Smithsonian Magazine*, June 18. <https://www.smithsonianmag.com/smart-news/nearly-2000-black-americans-were-lynched-during-reconstruction-180975120/>
- Franke, Katherine.** 2019. *Repair: Redeeming the Promise of Abolition*. Chicago: Haymarket Books.
- Franklin, Buck Colbert.** 1997. *My Life and an Era: The Autobiography of Buck Colbert Franklin*, edited by John Hope Franklin and John Whittington Franklin. Baton Rouge: LSU Press.
- Friedman, M.** 2019. "The Inflation Calculator." <https://westegg.com/inflation/>.
- Fussell, Melissa.** 2016. "Dead Men Bring No Claims: How Taking Claims Can Provide Redress for Real Property Owning Victims of Jim Crow Race Riots." *William and Mary Law Review* 57 (5) 1922–56.
- Georgetown Slavery Archive.** 1838. Articles of agreement between Thomas F. Milledy, of Georgetown, District of Columbia, of one part, and Jesse Beatty and Henry Johnson, of the State of Louisiana, of the other part. June 19, 1838. Box 40, File 10, Item 3a-h, Maryland Province Archives, Booth Family Center for Special Collections, Georgetown University.
- Gould, Elise.** 2020. *State of Working America Wages 2019*. Washington, D.C.: Economic Policy Institute.
- Haynes, George Edmund.** 1919. "For Action on Race Riot Peril." *New York Times*, October 5. <https://timesmachine.nytimes.com/timesmachine/1919/10/05/106999010.html?action=click&contentCollection=Archives&module=ArticleEndCTA®ion=ArchiveBody&pgtype=article&pageNumber=112>.
- Hirschman, Albert O.** 1958. *The Strategy of Economic Development*. New Haven: Yale University Press.
- Hudek, Stephen.** 2020. "Story of Ocoee Massacre Finally Being Told—100 Years After It Happened." *Orlando Sentinel*, October 29. <https://www.orlandosentinel.com/news/ocoee-massacre/os-ne-ocoee-what-happened-20201029-n3k7lqdbafg5re3qwxq2g3wnve-htm1story.html>.
- Katznelson, Ira.** 2005. *When Affirmative Action Was White: An Untold History of Racial Inequality in Twentieth-Century America*. New York: W.W. Norton.
- Kniesner, Thomas J., and W. Kip Viscusi.** 2019. "The Value of a Statistical Life." Vanderbilt Law Research Paper 19–15.
- Logan, Trevon and William Darity Jr.** 2020. "Look What Has Been Taken from Black Americans." *Bloomberg*, September 21. <https://www.bloomberg.com/opinion/articles/2020-09-21/calculating-reparations-look-what-s-been-taken-from-black-americans>.
- Levinson, Robert.** 2020. "Many Black World War II Veterans Were Denied Their G.I. Bill Benefits. Time to Fix That." *War on the Rocks*, September 11. <https://warontherocks.com/2020/09/many-black-world-war-ii-veterans-were-denied-their-gi-bill-benefits-time-to-fix-that/>.
- Marketti, James.** 1990. "Estimated Present Value of Income Diverted from Slavery." In *The Wealth of Races*:

- The Present Value of Benefits from Past Injustices*, edited by Robert F. America, 107–24. Westport CT: Greenwood Press.
- Max, Wendy, Dorothy P. Rice, Hai-Yen Sung, and Martha Michel.** *Valuing Human Life: Estimating the Present Value of Lifetime Earnings, 2000*. San Francisco: Center for Tobacco Control Research and Education.
- McKay, Tasseli.** Forthcoming. *Stolen Wealth, Hidden Power: The Case for Reparations for Mass Incarceration*. Berkeley: University of California Press.
- McLeod, Michael, and Joy Wallace Dickinson.** 2001. “Ocoee Race Riot Scars Generation.” *Orlando Sentinel*, February 6. <https://www.orlandosentinel.com/news/os-xpm-2001-02-05-0102050189-story.html>.
- Merrefield, Clark.** 2020. “The 1921 Tulsa Race Massacre and the Financial Fallout.” *Harvard Gazette*, June 18. <https://news.harvard.edu/gazette/story/2020/06/the-1921-tulsa-race-massacre-and-its-enduring-financial-fallout/>.
- Messer, Chris M., Thomas E. Shriver, and Alison E. Adams.** 2018. “The Destruction of Black Wall Street: Tulsa’s 1921 Riot and the Eradication of Accumulated Wealth.” *American Journal of Economics and Sociology* 77 (3–4): 789–819.
- Mitchell, Mary.** 2019. “A Century of Progress, But Still an ‘Invisible Divide’.” *Chicago Sun-Times*, July 26. <https://chicago.suntimes.com/columnists/2019/7/26/8931814/chicago-beaches-segregation-1919-riot>.
- Mittal, Anuradha, and Joan Powell.** 2000. “The Last Plantation” *Food First* 6 (1): 1–8.
- Mueller, Jennifer Catherine.** 2013. “The Social Reproduction of Systemic Racial Inequality.” Phd diss. Texas A&M University.
- National Humanities Center Resource Toolbox.** 2008. “On Buying One’s Freedom: Selections From 18th- and 19th-Century Slave Narratives.” *The Making of African American Identity: Volume 1, 1500–1865*, <http://nationalhumanitiescenter.org/pds/maai/emancipation/text1/buyingfreedom.pdf>.
- Neal, Larry.** 1990. “A Calculation and a Comparison of the Benefits of Slavery and an Analysis of Who Benefits.” In *The Wealth of Races: The Present Value of Benefits from Past Injustices*, edited by Richard F. America, 91–106. Westport: Greenwood Press.
- New York Daily Tribune.** 1865. “Newspaper Account of a Meeting between Black Religious Leaders and Union Military Authorities.” February 13, 1865. Freedmen and Southern Society Project. <http://www.freedmen.umd.edu/savmtg.htm>.
- Office of Program Policy Analysis and Government Accountability.** 2019. *Ocoee Election Day Violence—November 1920*. Tallahassee: Office of Program Policy Analysis and Government Accountability 19–15.
- Officer, Lawrence H., and Samuel H. Williamson.** 2022. “Annual Wages in the United States Unskilled Labor and Manufacturing Workers, 1774–Present.” Measuring Worth. <http://www.measuringworth.com/uswage/>.
- Ortiz, Paul.** 2006. *Emancipation Betrayed: The Hidden History of Black Organizing and White Violence in Florida from Reconstruction to the Bloody Election of 1920*. Berkeley: University of California Press.
- Perry, Andre, Jonathan Rothwell, and David Harshbarger.** 2018. *The Devaluation of Assets in Black Neighborhoods: The Case of Residential Property*. Washington, DC: Brookings Metropolitan Policy Program.
- Ransom, Roger L. and Richard Sutch.** 1990. “Who Pays Slavery?” In *The Wealth of Races: The Present Value of Benefits from Past Injustices*, edited by Richard F. America, 31–54. Westport CT: Greenwood Press.
- Ross, Jenna.** 2020. “The Racial Wealth Gap in America: Asset Types Held by Race.” *Visual Capitalist*, June 12. <https://www.visualcapitalist.com/racial-wealth-gap/>.
- Rothstein, Richard.** 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. New York: W.W. Norton.
- Rucker, Walter C., and James N. Upton, eds.** 2006. *Encyclopedia of American Race Riots*, Vol. 1 and 2, Westport: Greenwood Publishing.
- Schober, Daniel J., Bijou R. Hunt, Maureen R. Benjamins, Nazia S. Saiyed, Abigail Silva, Fernando G. De Maio, and Sharon M. Homan.** 2021. “Homicide Mortality Inequities in the 30 Biggest Cities in the U.S.” *American Journal of Preventive Medicine* 60 (3): 327–34.
- GBD.** 2019. Police Violence US Subnational Collaborators. 2021. “Fatal Police Violence by Race and State in the USA, 1980–2019: A Network Meta-Regression.” *Lancet* 398 (10307): 1239–1255.
- Shanks, Trina Williams.** 2000. “The Homestead Act: A Major Asset-Building Policy in American History.” Washington University Center for Social Development Working Paper 00-9.
- SoRelle, Ruth.** 2000. “Gap Between Death Rates for Blacks and Whites Remains as Large as in 1950.” *Circulation* 101 (12).

- Sources Select Resources.** “Red Summer of 1919.” SOURCES. https://www.sources.com/SSR/Docs/SSRW-Red_Summer_of_1919.htm.
- Stelzner, Mark, and Sven Beckert.** 2021. “The Contribution of Enslaved Workers to Output and Growth in the Antebellum United States.” Washington Center for Equitable Growth Working Paper 062421.
- Swinton, David H.** 1990. “Racial Inequality and Reparations.” In *The Wealth of Races: The Present Value of Benefits from Past Injustices*, edited by Richard F. America, 153–162. Westport CT: Greenwood Press.
- Temin, Peter.** 1999. “The Industrialization of New England: 1830–1880.” NBER Historical Paper 114.
- Thompson, Jeffrey.** 2019. “The GI Bill Should’ve Been Race Neutral, Politicos Made Sure It Wasn’t.” *Military Times*, November 9. <https://www.militarytimes.com/military-honor/salute-veterans/2019/11/10/the-gi-bill-shouldve-been-race-neutral-politicos-made-sure-it-wasnt/>.
- Turner, Sarah, and John Bound.** 2003. “Closing the Gap or Widening the Divide: The Effects of the G.I. Bill and World War II on the Educational Outcomes of Black Americans.” *Journal of Economic History* (63) 1: 145–77.
- Voogd, Jan.** 2008. *Race Riots and Resistance: The Red Summer of 1919*. Lausanne: Peter Lang.
- Webster, Ian.** 2021. “Stock Market Returns Since 1880,” S&P 500 Data. <https://www.officialdata.org/us/stocks/s-p-500/1880>.
- Wells-Barnett, Ida B.** 1920. *The Arkansas Race Riot*. Northern Illinois University Digital Library. <https://digital.lib.niu.edu/islandora/object/niu-gildedage%3A24320>.
- Williamson, Samuel H., and Louis P. Cain.** 2022. “Measuring Slavery in 2020 Dollars.” Measuring Worth. <https://www.measuringworth.com/slavery.php>.
- Wrigley-Field, Elizabeth.** 2020. “US Racial Inequality May Be as Deadly as COVID-19.” *Proceedings of the National Academy of Sciences* 117 (36): 28154–56.
- Yale, Aly J.** 2021. “What Is Real Estate Appreciation?” Millionacres, May 13. <https://www.millionacres.com/real-estate-market/what-is-real-estate-appreciation/>.
- Zinn Education Project.** 2022. “Nov. 2, 1920: The Ocoee Massacre.” Teaching People’s History. <https://www.zinnedproject.org/news/tdih/ocoe-massacre/>.

Slavery and the Rise of the Nineteenth-Century American Economy

Gavin Wright

Slavery is often named as the nation’s “original sin,” a term prompted by the glaring contradiction between the lofty Enlightenment rhetoric of the Declaration of Independence and the harsh reality of human bondage. Prevailing norms regarding slavery have of course changed across the centuries, but in this case the discrepancy was often noted and widely deplored by contemporaries, and not just among the anti-slavery founding fathers. For example, Revolutionary War hero Marquis de Lafayette is said to have declared a number of times in the 1790s: “I would never have drawn my sword in the cause of America, if I could have conceived that thereby I was founding a land of slavery” (as quoted by Clarkson 1846, p. 64). Yet the institution of slavery was embedded (though carefully not named) in the US Constitution of 1789 and persisted until ratification of the Thirteenth Amendment in 1865, in the wake of a cataclysmic Civil War.

In this essay, I focus on the debates over the role of slavery in the surge of nineteenth-century US growth. The United States was never poor by world standards, but in the 1780s, the country was fragmented, underdeveloped, and nearly stagnant economically. By 1860, per capita growth had accelerated to 1.7 percent per year, and the US economy was a formidable player, on its way to world leadership by the turn of the twentieth century. Did slavery play a primary and indispensable role in the rise of the US economy to world preeminence? This proposition has deep historical roots: Pro-slavery apologists argued that “slavery was the nursing mother of the prosperity of the North” (as quoted in Desmond 2019, an argument elaborated

■ *Gavin Wright is William Robertson Coe Professor of American Economic History Emeritus, Stanford University, Stanford, California. His email address is write@stanford.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.123>.

to great southern applause in Thomas Kettel's *Southern Wealth and Northern Profits* (1860). Karl Marx wrote in 1846 that "without slavery there would be no cotton, without cotton there would be no modern industry" (Marx 1963, p. 188). More recently, claims about the centrality of slavery to nineteenth-century US economic growth have featured in the national conversation on race, perhaps most notably in an intellectual insurgency known as the New History of Capitalism, some of which has filtered into the popular press through channels like the Pulitzer Prize-winning "1619 Project" published by the *New York Times* (2019).¹ Yet this proposition has been rejected by virtually every economic historian who has examined the issue.

This essay reviews more than six decades of scholarship on the economics of American slavery, with a focus on the interactions between slavery, economic growth, and development. I begin with a brief overview of New World enslavement of Native Americans, the transition to African American slavery, and the role of the British Empire in fostering slavery. I then discuss the slump of the US economy in the decades after 1776 and the surge of economic growth in the nineteenth century. One central issue is that much historical writing conflates *complicity* with slavery—for example, routinely engaging in business with slaveholders, including transactions that facilitated slaveholding and the domestic slave trade—with national economic impact. Complicity with slavery was indeed widespread in the free states. Indifference to the enslaved status of southern labor was more the rule than the exception among northern shippers, manufacturers, and financial firms. Further, such groups often supported slaveholder interests in national politics, fearful that turmoil over slavery would disrupt economic activity. However, accounts of the sources of US economic growth in the nineteenth-century suggest that slavery and the shift of the slave-owning South to cotton production early in the century had relatively little effect on growth for the nation as a whole. The southern economy shared in the growth acceleration, expanding lands committed to cotton production and raising labor productivity in the process. But the deeper sources of long-run US economic growth were improvements in technology, internal transportation, finance, and education, and the slave-owning South lagged in all of these areas. A simple summary of these patterns might be this: Slavery enriched slave-owners, but impoverished the southern region and did little to boost the US economy as a whole. To be clear, nothing in this argument should be taken to downplay the need for Americans to grapple openly with the nation's history of slavery and racial injustice.

Origins of North American Slavery

Coerced labor is as old as human history, but by the time of New World colonization, serfdom and other forms of bound labor had been in decline for centuries in western Europe. So why did nations that had long since abandoned slavery within

¹For critical overviews of the New History of Capitalism by economic historians, see Hilt (2017), Engerman (2017), Olmstead and Rhode (2018).

their own borders—taking their “free air” as a feature of national identity and pride—nonetheless preside over expansions of vast colonial empires based on slave labor? The contradiction was long a source of tension and commentary, but prior to the late eighteenth century, most observers found ways to rationalize the bipolar system by appeal to some combination of geographic necessity and national self-interest (Davis 1966; Drescher 1987).

To address this question, economic historians often draw upon Evsey Domar (1970), whose formulation was inspired by the re-enservment of free Russian peasants between 1550 and 1650, a response to outmigration into newly conquered territories in the east. Domar posited a trilemma with three basic elements: free land, free labor, and large-scale agriculture supporting a landlord class. His conclusion was that any two of these can coexist in a locality, but not all three at once. Free labor and large-scale agriculture describes western Europe of the 1700s, where land had become scarce enough to command rent from free tenants. Free land and free labor fits the US North, a region of family-sized farms. Free land and large-scale farming prevailed in the US South, but only under slavery. As to the choice between the two free-land alternatives, Domar declared that this was a political matter that he could not determine within his model.² Much is missing from this simple framework, particularly the violence required to establish and maintain slavery. But Domar’s insight, simplified though it is, helps to explain the re-emergence of slavery in New World colonies.

If slavery was going to exist, why enslave Africans rather than the native populations? Europeans did indeed enslave Indigenous Americans in North America, a practice going back to Christopher Columbus and predating significant importation of African slaves by more than a century. To be sure, slavery as a byproduct of Native American tribal warfare was ubiquitous before Columbus, but there seem to have been few “slaving societies,” waging warfare for the purpose of taking male captives for their labor. Native American slavery changed radically when the demand for slave labor from the arriving Europeans generated inter-tribal slaving wars similar to the African pattern. The largest mainland center of the Indian slave trade was Charleston, South Carolina, which exported tens of thousands of enslaved captives during the seventeenth century, predominantly to the sugar islands of the Caribbean, but also to the tobacco colonies of the Chesapeake. The Charleston hub destabilized Native American tribal relations throughout the southeast by setting in motion an active trade in Indian slaves. As a starting point for further discussions of Native American slavery, interested readers might begin with Gally (2002) and the collection of essays on Indian slavery in various regions of North America in Gally (2009).

Indian slavery declined in North America in the eighteenth century largely for reasons of demography. Indigenous populations fell precipitously, primarily because of exposure to new diseases such as smallpox, influenza, and malaria,

²As Domar acknowledged, the association between slavery and “free” land or resources had a long intellectual history. A prominent advocate was Nieboer (1900).

but also through enslavement and brutal work regimes. Diplomatic and political factors were at work as well. From the beginning, Indigenous and African slavery were viewed differently by European countries; for example, the Spanish monarchy banned Indian slavery in 1542. This act was not enforceable, but the illegality of the practice limited its public visibility and geographic scope (Reséndez 2016). In South Carolina, proprietors also discouraged Indian slavery, expecting a greater return on their investments from settled plantation agriculture than from slave sales, whose profits were largely captured by military commanders. By the eighteenth century, slavery in the eastern British colonies of North America that became the United States was closely identified with African laborers.³

African slavery was not invented in British America: indeed, the English were latecomers both to colonization and to slavery. When the settlers of Barbados determined that their most profitable cash crop was sugar, and then turned in the 1640s from indentured servants to African slaves, they were buying into what was already a well-established transatlantic slave trade. The largest buyers of that time were Spanish and Portuguese colonies, and the largest slave-traders were the Dutch, who were only too happy to extend their sales to the new English colonies. When the Royal Africa Company of England was chartered in 1672, British America acquired its own direct supplier of slaves. The slave trade was opened to competition in 1689, and expanded supply channels help to explain the transition from indentured servitude to slavery in the tobacco colonies of the Chesapeake between 1690 and 1710. A deeper force for the change was that improved mortality conditions tipped the cost calculus in favor of slaves-for-life over indentured servants for a limited term of years. Also by that time, the legal distinction between the two categories of bound labor—unclear for several decades after the first arrival of Africans in 1619—had been clarified, as well as their respective racial identities (Menard 2001; Galenson 1981, pp. 141–57).

The fact that African slaves were obtainable only at prices set in competition with the rich sugar islands of the Caribbean helps to explain the geographic pattern that puzzled Domar (1970): Why did northern farmers not make more use of slave labor? It was not a matter of ideology. Nor was slavery unsuited to grain farming or mixed husbandry, as is often asserted; there are prominent examples of deployment of slave labor on commercial wheat farms in both the eighteenth and nineteenth centuries (Irwin 1988; Heerman 2018, pp. 18–37). New Englanders complained as much as landowners anywhere about the cost of employing free labor: for example, a Boston Puritan named Emmanuel Downing wrote in a 1645 letter to Massachusetts Governor John Winthrop that “our servants desire to plant for themselves, and not stay except for verie great wages,” just as Domar posited (Winthrop 1947, pp. 38–39). Downing favored war with the Narragansett tribe to

³Indian slavery persisted much longer in the western and southwestern regions of North America. Reséndez (2016) reports that American explorers entering the Great Basin in the 1830s were surprised to discover an active trade in Indian slaves, often exchanging horses for children destined to purchasers in New Mexico and California.

Table 1
Population of British America by Region, 1780

	<i>Total</i>	<i>Black</i>	<i>Black (%)</i>
New England	712,829	14,427	2.0
Middle Atlantic	466,973	39,369	8.4
Upper South	1,156,481	406,278	35.1
Lower South	251,071	127,831	50.9
British West Indies	570,000	491,000	86.1

Source: Historical Statistics of the United States Millennial Edition (volume 5, tables Eg 1–59).

Note: New England: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont. Middle Atlantic: New Jersey, New York, Pennsylvania. Upper South: Delaware, Kentucky, Maryland, North Carolina, Tennessee, Virginia. Lower South: Florida, Georgia, South Carolina.

Table 2
Population of the United States by Region, 1790

	<i>Total</i>	<i>Slave</i>	<i>Free Black</i>	<i>% Black</i>	<i>% Black Pop. Free</i>
New England	1,009,408	3,763	16,911	2.0	81.8
Middle Atlantic	958,632	36,323	13,855	5.2	27.7
Upper South	1,630,807	521,200	30,324	33.8	5.5
Lower South	331,621	136,358	2,194	41.8	1.6

Source: Historical Statistics of the United States Millennial Edition. Volume 2. Table Bb 1–98; Volume 1. Table Aa 36–92; Table Aa 5607–5707.

Note: Florida is not included.

capture “woeman and Children enough to exchange for Moores, which will be more gaynefull pillidge for us than we conceive...” The problem with this plan was that in most northern locations, the marginal revenue product of farm labor was not high enough to warrant the high international price of African slaves. But this economic marginality does not imply that these areas were unconnected to the larger slave-based Atlantic economy.

In the eighteenth century, slavery was legally recognized in all parts of British America, as suggested by the ethnic breakdown of the British American population in 1780 shown in Table 1. In the mainland colonies that would become the United States, although the slave population was concentrated in the tobacco regions of the Upper South (as shown in Table 2), the institution was also significant in fertile areas of the Middle Atlantic, and in New York City, where more than 40 percent of the households held at least one slave in the colonial period (Foote 2004, p. 12).

Table 3

Imports into England*(in thousands of British pounds)*

	1699–1701	1722–1724	1752–1754	1772–1774	1784–1786	1794–1796
Sugar	630	928	1,302	2,364	2,614	5,572
Tobacco	249	263	560	519	536	368
Rice		52	167	340	155	178
Indigo	85	152	97	167	127	124

Source: Davis (1962, pp. 300–301); Davis (1979, pp. 110–114).

Colonial British North America and Slave-Based Commerce

The British Atlantic economy broadly understood flourished between 1700 and 1775, but the mainland North America colonies were marginal contributors to this historic phenomenon. The largest mainland export was tobacco, corresponding to the main concentration of slaves in Maryland and Virginia. But the tobacco trade, measured by imports by England, paled in comparison with sugar exports from the West Indies, which were more than 4.5 times as large by the end of the nineteenth century, as shown in Table 3. Rice exports from Georgia and South Carolina (whose production was more fully based on slavery) were even smaller, though growing. Notably, cotton does not appear on the list; that crop was virtually nonexistent on the mainland in colonial times. As might be expected from the quantitative dominance of sugar and the smaller white share of the island populations, the fortunes of tobacco planters were far outstripped by the “prodigious riches” of sugar planters, the true plutocrats of that era. The value of physical wealth in the West Indies was more than ten times greater than that of the southern mainland colonies (Burnard 2001, p. 520).

The British Empire constituted a complex interdependent economic system, protected against outsiders by military superiority (Findlay 1990; Darity 1982). Britain’s northern colonies, as imperial insiders, were beneficiaries of this regime. New England and the Middle Atlantic had relatively few slaves, but the British Caribbean was their largest trading partner. To feed slave labor in that region, the Middle Atlantic shipped grain products, beef, and pork, while New England sent fish, whale products, and lumber (McCusker and Menard 1985, pp. 108, 199). In addition, during 1768–1772, two-thirds of New England’s “invisible” earnings—shipping services, warehousing, shipbuilding, and finance—arose from Caribbean commerce, providing the region with its largest single overseas revenue source (Richardson 1991). The New Historians of Capitalism sometimes imply that these connections to slavery are only now being discovered, but they have been known for generations. The Black historian Lorenzo Greene (1942, pp. 68–9) wrote: “On the eve of the American Revolution [the slave trade] formed the very basis of the economic life of New England. The vast sugar, molasses, and rum trade, shipbuilding, the distilleries,

a great many of the fisheries, the employment of artisans and seamen, even agriculture—all were dependent upon the slave traffic.”

The question before us is whether slavery’s contribution to the colonial economy can be said to constitute the basis for accelerated growth in the new nation. Certainly, there were lasting legacies of the slave trade that might be identified as “infrastructure” in the broadest sense. The growth of cities is sometimes taken as an index of development, and the United States had a handful of ports that owed much to slave-based commerce. Rich merchants in these cities sometimes used their wealth to found universities, some of which no doubt provided economic benefits in subsequent centuries. Wilder (2013, p. 17) shows that the first five colleges in British America were major beneficiaries of the slave trade and slavery. Wilder (p. 30) adds: “Harvard then became the first in a long line of North American schools to target wealthy planters as a source of enrollments and income.” A number of early-American libraries can trace their origins to slave profits (Moore 2019). Among the American entities that can be fairly indicted for trying to forget or bury their slave pasts are the northern states that abolished slavery after the American Revolution, and their constituent cities and universities.

Yet despite these earlier connections and complicities, it would be difficult to make the case that accelerated US growth in the first half of the nineteenth century grew out of this slave-based past in any direct or fundamental way. The United States was already an affluent nation on the eve of the American Revolution, certainly among the richest in the world in per capita income (Lindert and Williamson 2016). But the high living standards of that time reflected a favorable disease environment and cheap foodstuffs, not rapid growth per capita. The US economy in 1776 was affluent by world standards, yet deeply underdeveloped. It had no banks or domestic capital markets, almost no internal transportation, no modern mining or manufacturing activity, and weak institutions for the advancement of useful knowledge. Most economic advancement during the colonial era was *extensive*—that is, growth in population and in land area—with little productivity increase within sectors (Rosenbloom 2018). There were gains from urbanization and commerce, but the trade structures supporting these were devastated by the Revolutionary War and then by exclusion from the British Empire. For the northern half of the country in particular, the path to accelerated economic growth in the nineteenth century did not lie in reconstructing the long-distance trade networks of the colonial era.

Comparing Tables 1 and 2 points toward a distinctive feature of mainland slavery. Even during a decade when markets were shriveling and prices falling, the slave population grew, mainly because of favorable epidemiological conditions and cheap food. Over the entire history of the transatlantic slave trade, it is estimated that no more than 5 percent of coerced African migrants came to mainland North America (Eltis and Richardson 2008, pp. 48–51).⁴ Yet this peripheral fringe of the

⁴The 5 percent figure refers to the period ending in 1807, when the United States closed the African slave trade. For the entire period through 1867, the North American share would be 3.6 percent.

trans-Atlantic slave trade had reached a total of four million slaves by 1860, roughly half of the slave population of the Western Hemisphere at that time.

The American Revolution and the Eventual Acceleration of Economic Growth

It has become popular to assert that slave-owners joined the American Revolution in order to protect slavery, having been spooked by the 1772 decision of the English Court of King's bench in *Somerset v. Stewart*, which held that under English common law, an enslaved person could not be forcibly taken for sale in Jamaica, thus denying the legality of slavery in England under the common law. There is controversy over how much attention southern slave-owners actually paid to that decision and its relatively limited holding. But if the southern slave-owners did join the revolution for economic reasons, their choice was a major miscalculation. The US Revolution and its aftermath inflicted massive damage to the mainland economies, the greatest hits taken by the slaveholding colonies of the South. Lindert and Williamson (2016) estimate that per capita personal income fell by 30 percent in the South Atlantic region between 1774 and 1800, while declines in New England and the Middle Atlantic were much smaller. The authors attribute the southern fall to disruption of pre-war imperial export markets as well as wartime destruction, including the loss of slave labor.

The US economy did eventually recover, and per capita growth accelerated after 1790, reaching roughly 1.7 percent per year by the mid-nineteenth century.⁵ The timing of the rise in growth rates cannot be determined with precision, both because statistical sources are limited and because the Napoleonic Wars of 1793–1815 (which can be taken to include the US War of 1812 against Britain) created extraordinary external conditions, making it all but impossible to distinguish trends from shocks. Even so, the consensus among economic historians is that long-term growth processes were underway in the 1790s, if not earlier. Explanations for growth acceleration at this time typically give prime place to certain institutions established by the US Constitution of 1789 and the fiscal reforms of the first Washington administration undertaken by Alexander Hamilton. Together, these structures brought fiscal credibility to the central government, stabilized the currency, and channeled policies of the states in growth-promoting directions (Irwin and Sylla 2011). Within this institutional context, some analysts assign primacy to developments in financial markets (Rousseau and Sylla 2005), while others see transportation investments, typically government-sponsored, as the driving force (Wallis and Weingast 2018). Because these two developments were complementary, it may not be necessary to choose between them. Buoyed by expanding opportunities, American farmers rapidly migrated west and commercialized, adopting mechanized equipment by

⁵ Alternative estimates are presented and discussed in Rhode and Sutch (2006, especially pp. 15–19).

the 1850s (Weiss 1993). Responding to growing domestic markets and protected by tariffs, American manufacturing industries deployed innovative technologies to generate productivity growth rates averaging 2.5 percent per year between 1820 and 1860 (Sokoloff 1986).

Among the institutional changes established in the wake of the US Revolutionary War are the abolitions of slavery in the northern states, which were accompanied by a new perception of the relationship between slavery and economic progress. Beginning with Vermont in 1777, the northeastern states enacted gradual abolition laws, ending with New York in 1799 and New Jersey in 1804. These acts were unintended consequences of the Revolution, impelled by the “excruciatingly conspicuous” divergence between national ideals and the reality of slavery (Wood 1992, p. 186). The campaigns were often led by free Blacks, who brandished the Declaration of Independence as an “abolitionist tract” (Quarles 1976). One may suspect that these abolitions had little economic significance, because slaves were not a large share of the population in any of these states. But slave-owners in New York and New Jersey fought hard to retain their slaves, and although they succeeded in making emancipation across this group of states extremely gradual, the direction was clear. By contrast, slavery was not abolished in the British Empire until 1833.

Arguments over slavery in what were called the “Northwest Territories”—what would become Ohio, Indiana, Illinois, Michigan, Wisconsin, and part of Minnesota—were often explicitly couched in terms of economic development. Article VI of the Northwest Ordinance of 1787, possibly the most significant institutional manifestation of postwar antislavery feeling, excluded slavery from these territories. The origins of Article VI are obscure. Its enactment came as a shock to the many slaveholding settlers in Ohio and Indiana, who inundated Congress with petitions for repeal. Only the threat of congressional veto induced Indiana to enter the Union as a free state in 1816 (Rosenberg and McClurg 1968). Illinois had an even longer tradition of slaveholding and pro-slavery politics (Heerman 2018). Here too, Illinois joined the new country as a free state in 1818 only because of the Ordinance, but in 1824 the state political leadership proposed a new constitutional convention, which was widely understood as a referendum on slavery. Advocates rehearsed the older view that slavery was necessary for productive settlement in remote areas: “Look at those trains of wagons and their splendid teams, their carriages and their gangs of negroes. They are going to fill up Missouri, and make it rich, while our State will stand still or dwindle, because you won’t let them keep their slaves here” (quoted in Guasco 2001, p. 21).

The debate had elements of regional culture clash, but also at work was a sense of linkage between slavery and migration, and between slavery and land values. In each of the Northwest territories, concerns that slavery was needed for productive development were overtaken by an influx of settlers from the northeastern states. Rev. Thomas Lippincott, corresponding secretary for the “Madison Association to oppose the introduction of Slavery in Illinois,”

pointed out that population and land values increased more rapidly in Pennsylvania than in Virginia, concluding that “the existence of slavery in one, and its non-existence in the other state, has caused the discrepancy” (quoted in Guasco 2001, p. 23).

Slavery and the Geography of Growth

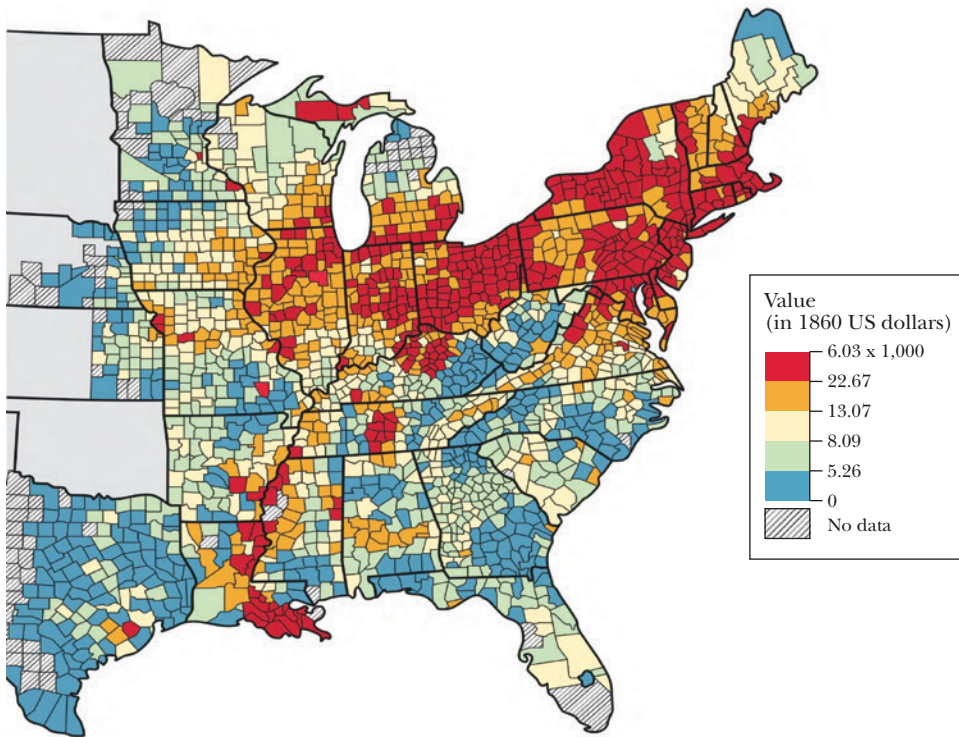
The slave South shared in the growth acceleration of the late eighteenth century, coming out of its economic doldrums in the 1790s. Popular history usually links this take-off with Eli Whitney’s invention of the cotton gin in 1793, but historians have long understood that this event was a blip in a more extended transition (Chaplin 1991; Lakwete 2003). As rising British production of cotton goods put pressure on traditional supply sources, global prices for raw cotton surged in the 1780s and 1790s (Broadberry and Gupta 2009, p. 290). During these decades, southern planters were actively searching for an alternative to tobacco, as well as to grains and indigo, whose prices were falling. After much experimentation, they found a likely answer in upland cotton, a discovery that in turn created the production bottleneck addressed by the introduction of the cotton gin. The rise of cotton did breathe new economic life into slavery, but it was not obvious at the outset that the crop was particularly well suited to coerced labor. Early promoters pointed to the minimal capital outlay required for cotton, predicting “leveling effects” from its spread through rural areas, enabling “the poor” to be “elevated to this middle grade of society” (Klein 1990, pp. 248–9). Cotton became a slave crop, not because of technological imperatives, but because the farmers looking for alternatives were already slave-owners. As historian Joyce Chaplin (1991, p. 199) writes: “Early cotton cultivators used cotton to preserve a world already shaped by commercial agriculture and slavery.”

Although the southern economy also grew in the first half of the nineteenth century, the geography of its expansion was markedly different. Whereas farmers in the free states moved steadily westward, responding to (and demanding) new transportation facilities, slave-owners leapfrogged to the rich cotton lands of the southwest. The movability of slave property meant that these “laborlords” had no need to recruit workers to help with land clearance and no reason to entice settlers with innovative financial terms. Beginning in the 1790s, turnpikes built by state-chartered corporations crisscrossed the northeastern states, but (as Tocqueville noted) southern states were barely represented in this activity.⁶ The next wave of transportation infrastructure, the canal boom of 1815–40, was also overwhelmingly regional; five times as many miles of canals were constructed in northern than in southern states. Some railroads came southward during this period, but railroad

⁶Klein, Daniel B., and John Majewski. 2008. “Turnpikes and Toll Roads in Nineteenth-Century America.” <http://eh.net/encyclopedia/turnpikes-and-toll-roads-in-nineteenth-century-america/> (accessed June 17, 2021).

Figure 1

Farm Value per Acre of Land on Farms, by County, 1860



Source: Wright (2006, p. 65).

Note: Colorization by Mary Lee Eggart is gratefully acknowledged.

mileage density was more than three times greater in the northern states; southern lines were said to be “generally inferior in construction, rail, motive power and rolling stock” (Wright 1986, pp. 21–24). Testimony to the South’s economic isolation is the fact that the region maintained a railway gauge standard separate from that of the rest of the country; it conformed to the national standard only in 1886 (Gross 2016).

These patterns are reflected in the spatial pattern of farm value per acre shown in Figure 1. Overall, there is a stark North-South contrast in the value of farmland. More specifically, land values in the free states of the western migration of that time—like Ohio, Indiana, and Illinois—are much higher than the values of the southwestern states—like Alabama, Mississippi, and Tennessee—except for the most fertile riverine areas. The visual impression of Figure 1 is confirmed more rigorously by Bleakley and Rhode (2021), who estimate the effects of the free-slave boundary on economic outcomes, like the Ohio River boundary between the free

state of Ohio and the slave state of Kentucky. Even after controlling for longitude, climate, and detailed soil variables, the authors find a substantial negative impact of slave-state location on population density and farmland values.

To be clear, economic growth in the slave South was real. It can be misleading, however, to compare regional economic performance using current income estimates alone, without also considering the sources of growth and their sustainability. First, the population shift from the low-productivity southeast to the high-productivity southwest was a substantial contributor to regional growth, but it was inherently transitional and limited. A second contributor to productivity growth, documented by Olmstead and Rhode (2008), is improvement in cotton seeds. This source is less obviously self-limiting. But Olmstead and Rhode (2011) argue that improved cotton seeds constituted a “regionally biased innovation,” in that they had greater impact on the already-superior soils of the southwest. Olmstead and Rhode’s evidence from plantation records shows remarkable gains in picking productivity during the pre-Civil War era. However, these growth rates were clearly declining by the 1850s, as one would expect from a growth source with a strong regional-shift component. A third contributor to southern antebellum growth, which stands as background-context to all of the supply-side factors, is persistent growth of world demand for US cotton, averaging more than 5 percent per year for more than 60 years. This growth source too did not prove sustainable across the remainder of the nineteenth century; growth of cotton demand fell to an estimated 1.5 percent per year between 1866 and 1895 (Wright 1974). In summary, there is good reason to believe that economic growth in the slave South would have slowed after 1860, even in the absence of a devastating war and radical disruption of the region’s labor system.

If the medium-term sources of growth in the pre-Civil War slave economy of the South were not sustainable, then per capita income estimates for two or three census years—especially estimates related to a year of peak cotton demand and extraordinary yields like 1860—are not a reliable way to assess regional progress. A better approach is through the accumulation of wealth. Table 4 reports total and per capita levels of real and personal property from the censuses of 1850 and 1860, building on estimates of slave values constructed by Ransom and Sutch (1988). It is immediately apparent that the question of regional wealth accumulation turns on the treatment of slave values, which accounted for nearly half of all tangible southern wealth in 1860. If slave values are included, the South was the wealthiest region per capita; if only the free population is counted (about two-thirds of the total), the southern advantage was immense, more than 50 percent in 1850, 80 percent in 1860. But if instead we only consider nonhuman wealth, as we would for a modern economy, the North had the advantage, by 80 percent in 1850 and 64 percent in 1860. The northern regional edge prevailed even for the free population alone, though these differences were smaller.

What economic mechanism or process was at work to generate this pattern? Ransom and Sutch (1988) favor a macroeconomic model in which the rise of slave values satisfied the desire to accumulate wealth, and in this way “crowding

Table 4
Regional Wealth in 1850 and 1860

	1850		1860	
	North	South	North	South
Physical wealth (billions of dollars)	\$4,474	\$2,844	\$9,786	\$6,332
Value of slaves (billions of dollars)		\$1,286		\$3,059
Nonslave wealth (billions of dollars)	\$4,474	\$1,559	\$9,786	\$3,273
Wealth per capita	\$315	\$316	\$482	\$569
Nonslave wealth per capita	\$315	\$174	\$482	\$294
Nonslave wealth per free capita	\$315	\$266	\$482	\$449
Wealth per free capita	\$315	\$483	\$482	\$868

Sources: *Preliminary Report of the Eighth Census* (1862, 194–195); Ransom and Sutch (1988, 150–51). Free Black population figures are from the US Census Bureau (1918, 57).

out” real capital formation, such as land-clearing, fencing, or farm equipment. The argument is strengthened when blended with a complementary analysis of the relationship between slave ownership and incentives to invest in “development” of localities. Because slaves were movable personal property in a well-developed regional market, their value was virtually independent of local development. Owners could borrow across long distances, reducing the need to cultivate local credit relationships, as was common in the free states. Because slaves provided captive labor for setup tasks like land-clearing, owners had little reason to engage in recruitment of workers or settlers, activities that engaged extensive entrepreneurial energies in the states where slavery was prohibited.

Thus, although northern travelers such as William Seward or Frederick Law Olmstead provided vivid eyewitness accounts of southern economic backwardness, slave-owners themselves had little reason to regret their choice to retain their peculiar institution. Indeed, the more intellectually inclined sought to make backwardness a positive virtue. As Confederate politician Louis Wigfall told a British correspondent in 1861 (as quoted in King 1970, p. 126):

We are a peculiar people, sir! ... We are an agricultural people; we are a primitive but civilized people. We have no cities—we don’t want them. We have no literature—we don’t need any yet ... We want no manufactures: we desire no trading, no mechanical or manufacturing classes ... As long as we have our rice, our sugar, our tobacco, and our cotton, we can command wealth to purchase all we want from those nations with which we are in amity, and to lay up money besides.

In many respects, the contrast in tangible wealth between slave and free regions shown in Figure 4 understates the difference across regions. By measuring only tangible wealth, the table neglects education, a sector in which the slave South

lagged badly. Slave-owners as a class had obvious reasons to fear slave literacy, so teaching a slave to read was prohibited in most southern states. Slave-owners also had little reason to pay for schools for the free population, whom they did not plan to recruit as settlers and did not expect to hire as employees. Only 35 percent of the region's school-age white population were enrolled in 1860, compared to 72 percent elsewhere—where the school year was 70 percent longer. Historian Keri Leigh Merritt (2017) estimates that more than one-third of the free population were “poor whites,” owning neither land nor slaves. Poor whites were subject to vagrancy laws, jail for debt, and even having their children “bound out” as indentures. Merritt is careful not to equate poor whites with slaves. But clearly a large segment of the white southern population did not derive economic benefits from the slave regime.

Complicity versus Economic Impact

In light of this record, it should not be surprising that, over the past two centuries or more, the impoverishment, limited education, and poor health of southerners has exerted a drag on US economic performance. Yet the idea persists, and now seems to enjoy wide circulation, that the slave South somehow played a pivotal role in launching the United States on the path towards prosperity. In my view, the essence of the matter is a failure to distinguish “complicity”—routinely transacting with slave-owners, or buying and selling products of slave labor—with an impact on growth deriving from slavery as a labor system and form of property.

The conflation is illustrated in a book by three journalists with the informative title *Complicity*, identifying numerous instances of northern engagement with slavery, from the New England cotton textile industry to New York merchants and shippers, insurance companies, and New Haven carriage makers who catered to the tastes of wealthy planters (Farrow, Lang, and Frank 2005). The cumulative impact of the examples is powerful, but the book makes no effort to describe the national growth process in order to assess slavery's contribution to it. This limitation does not deter the authors from concluding that: “It is obvious that, at the very least, America's extraordinary ascent into the world arena would have taken far longer than it did [had there been no slavery]” (p. 215). The methodology seems to be: If people made money doing something bad, then these activities must have been advantageous for the nation as a whole.

Complicity at least has the virtue of stating a counterfactual—a nation without slavery—though the authors make no effort to explore what such a world would have looked like. One underlying assumption is that cotton required slavery. This claimed association is explicit in Beckert (2014, p. 110):

Cotton demanded quite literally a hunt for labor and a perpetual struggle for its control. Slave traders, slave pens, slave auctions, and the attendant physical

and psychological violence of holding millions in bondage were of central importance to the expansion of cotton production in the United States and of the Industrial Revolution in Great Britain.

It is far from clear, however, that the antebellum South produced more cotton than the region would have, had slavery been abolished at the time of adoption of the US Constitution. Yes, slave-owners moved quickly to the rich soils of the cotton frontier, and yes, they drove their enslaved laborers brutally. But to maintain the value of their property, they also actively supported the closure of the African slave trade in 1807 (Mason 2000). Enslaved Africans might have been replaced by free migrants, but slave-owners do not recruit free settlers or workers to the South. Further, they neglected transportation infrastructure, which was central to the rapid westward migration of family farmers into the Northwest Territory.

Is there any good reason why the cotton South might not have developed along family-farm lines similar to the North in the absence of slavery? True, slaveless southern farms produced relatively small amounts of cotton in the antebellum era, when marketing and credit facilities were poorly developed. But these types of services could have emerged rapidly in response to demand, as they did in the free states. This counterfactual is developed in more detail in Wright (2020, pp. 369–373). What we do know is that small southern farmers moved into cotton production quickly after the Civil War, bringing cotton prices down to pre-war levels within a decade after the surrender of the Confederate forces at Appomattox in 1865.

Economic historians now routinely demand that counterfactuals be specified clearly, as part of any discussion of proposed historical effects of *x* upon *y*. In contrast, historians often object to counterfactuals, saying they prefer to write about the history that actually happened rather than hypothetical histories that might have been. In deference to this perspective, let us set aside the question whether cotton production really needed slavery, and focus instead on the role that cotton may have played in the US growth acceleration in the decades before the Civil War.

The most direct channel is through the cotton textiles industry, an early manufacturing success. Complicity here is hard to deny: New England mills used slave-grown cotton, generally without qualms or protest. But was cotton vital for the US industrial revolution? Cotton textiles was the largest antebellum manufacturing industry, but it was only one among many, accounting for less than 10 percent of employment and 6 percent of value-added in 1860. Other manufacturing industries showing rapid productivity growth before 1850, such as furniture, glass, hats, paper, tanning and coaches/harnesses, had little if any connection to cotton or slavery (Sokoloff 1986).⁷

⁷The textile machinery industry was indeed an important vehicle for the diffusion of machine technologies in the nineteenth century (Rosenberg 1963; Lozier 1986). But to the extent that this upstream activity was a function of the scale of the parent industry, it owed more to the protective tariff than to slavery.

Even for cotton textiles, a causal association based on country-of-origin of the raw cotton supply makes little economic sense. Raw cotton is a bulky, light-weight commodity that travels easily, so that transportation costs have little effect on industry location. True, some cotton textile industry leaders cultivated close personal ties with southern planters and favored (along with the majority of Americans) compromise on slavery-related political issues—like their British counterparts, they were fearful of violent disruption of cotton supplies. But even these Cotton Whigs generally took anti-slavery stances like opposition to the annexation of Texas in 1845 (which entered the United States as a slave state) and the extension of slavery into Kansas in 1854 (O'Connor 1968; Abbott 1991). Of far greater significance to leaders of US industry was the series of tariffs against imports of manufactured goods, starting in 1816, rising in 1824, and rising still higher in 1828. Such tariffs were essential for survival of most domestic cotton textile manufacturers (Harley 1992, 2001). However, they were bitterly opposed by the South, on the ground that they benefited Northern cotton-buyers at the expense of southern cotton-sellers. The tariff issue nearly split the nation during the Nullification Crisis of 1832–33. Thus, it seems absurd to add cotton textiles to cotton production as a measure of the economic impact of southern slavery on the national economy.

Another proposed claim that “cotton drove US growth” is based on timing: the argument goes that an early spurt in cotton exports triggered southern demand for western foodstuffs and for northeastern manufactured goods, thereby launching these regions on their modern growth paths. This version revives a tradition articulated by Douglass North in 1961. Though he generally said little about slavery, North (1966, pp. 66–7) wrote: “Cotton was strategic because it was the major independent variable in the interdependent structure of internal and international trade. The demand for western foodstuffs and northeastern services and manufactures were basically dependent on the incomes received from the cotton trade.” However, decades of cliometric research has considered and rejected each major part of this argument.

Fishlow (1965) and Lindstrom (1970) show that the South provided no more than a limited market for western foodstuffs. For example, Lindstrom (p. 113) concludes: “[T]he needs of the Lower South for flour and corn were insufficient to absorb the output of these products from the upper South, to say nothing of their serving as a major outlet for western produce.” In other words, the northeast was the nation’s major food deficit region, even during the early years of the nineteenth century when what were then the “northwest” states, like Ohio, Indiana, and Illinois, marketed most of their goods downriver through New Orleans. In the 1860 Census, census “marshalls” went from farm to farm, recording dozens of pieces of detailed information. The landmark Parker-Gallman sample collected this data from manuscripts for 5,229 farms from the 1860 census, specifically to test North’s interregional trade hypothesis (Parker and Gallman 1992). Drawing on the sample, Gallman (1970) showed that most cotton plantations were themselves self-sufficient in food, planting ample corn crops to spread the fixed costs of slave labor across the year.

As a market for manufactured goods, the South was never dominant, and its share diminished over time as its share of national income declined. Applying capture-recapture techniques (borrowed from the study of wildlife populations) to analyze the coastal trade, Herbst (1978) estimates that no more than 16.4 percent of northern manufacturing output went to southern ports in 1839, and the share directly attributed to cotton income was less than 4 percent. In her study of manufacturing in the Philadelphia region, Lindstrom (1978) finds that goods rarely sold long-distance before 1840, and when they did, these markets were usually in the east. Longer-distance trade grew over time, but mainly along east-west lines, following the paths of migration and transportation.

What about patterns of extra-regional transactions involving financial markets and services that go beyond the purchase of slave-made products in impersonal markets, or supplying luxury products to slave-owners? Ships based in northern cities participated actively in the coastal slave trade in the 1810s and 1820s, as part of the “cotton triangle” trade connecting New Orleans and Liverpool (Schermerhorn 2016).⁸ New York steamship companies vigorously competed for the Texas slave trade (Schermerhorn 2015, chapter 7). Northern banks and joint stock companies responded eagerly to opportunities opened by dispossession of Cherokees, Chickasaws, and Choctaws from prime cotton lands in the 1830s (Saunt 2019). During this boom decade, the Natchez (Mississippi) branch of the (second) Bank of the United States under Nicholas Biddle offered accommodation loans to planters so lavishly that the bank found itself in possession of numerous slaves and several plantations after the financial collapse of 1839 (Kilbourne 2006). In these cases and others, northern interests were not just trading with slave-owners, but investing in and facilitating the operation of the slave system itself.

The question before us in this essay is whether these financial and trade interactions with the slave South initiated, accelerated, or shaped nineteenth-century financial and economic development in the northern states to any substantial degree. Such connections are not obvious. In the aftermath of the Revolutionary War, capital markets in northeastern US cities were liberated from British mercantilism and flourished before cotton even appeared on the US scene. The main traded assets were US bonds and shares of state-chartered banks, insurance companies, and turnpike and bridge corporations (Sylla 1998). New York City was both the nation’s leading port city and financial center before cotton became economically significant. Because the city is one of the world’s largest natural harbors, it is more accurate to say that “New York attracted cotton” than that “cotton made New York.” Notably, even as the “cotton triangle” trade route took shape after 1815, the eyes of the city fathers were on trade with the western free states. For example, the Erie Canal was approved by the state legislature in 1817—a state project only because of

⁸Schermerhorn writes that the “cotton triangle” was actually an “irregular polygon” (2015, p. 44). The expression “cotton triangle” is reminiscent of, but distinct from, the triangular trade of the eighteenth century, which carried slaves from Africa to the Americas, primary products from the New World to Britain, and British manufactured goods to Africa and the Americas.

southern opposition to federal funding—at the urging of Governor DeWitt Clinton, who had been mayor of New York City and a member of the Erie Canal Commission from 1810 to 1824. The marketing of canal bonds was as much a stimulus to New York’s financial development as the cotton trade, while the canal itself confirmed and extended the flow of domestic commerce along east-west lines (Myers 1931; Olmstead 1972). Transatlantic shipping was not the direction of the future for the US economy.

It is quite true that flows of “outside” capital into the South were significant, both from northern and European investors. In recent years, economic historians have come to a greater appreciation of the financial features of the slave regime. Slave values served as backing for extensive networks of credit, both local and long-distance, because slave property was mobile and hence liquid (Kilbourne 1995; Martin 2010, 2016; González, Marshall, and Naidu 2017). It is thus plausible that the attractiveness of slave property as collateral encouraged greater investment in the Southern region than would otherwise have occurred. To the extent that these flows facilitated the westward movement onto superior cotton lands, one may say that they contributed to faster economic growth. The best estimates, however, suggest that an even larger volume of foreign investment went to the free states, in support of canal and railroad construction (Wilkins 1989, pp. 53–72). During the same era, “old money” capitalists in the northeast also saw investment prospects much more favorably in the West than in the South (Johnson and Supple 1967). Long-distance investments in both often ended in default. However, a central difference was that infrastructure projects left a legacy of physical capital and a revolutionized transportation system; investments in the slave South promoted geographic relocation, but mainly served to drive up the price of slaves, the constraining factor.

Slavery, Regional Conflict, and Civil War

If one holds the belief that slavery tied together the North and the South in a supposedly complementary and mutually advantageous pro-growth relationship, it becomes difficult to understand why the two regions waged unrestrained warfare against each other for four years in the 1860s. Slavery was unquestionably the central cause of the US Civil War, and no reputable historian believes otherwise. But slavery was more than a moral or ideological issue; it interacted with developments of the time, and with pressing political and economic debates. In the lead-up to the US Civil War, economic and political leaders in both North and South did not see the emerging trends as mutually beneficial or even compatible.

Probably the greatest concentration of self-interested proslavery sentiment in the free states was New York City, also the site of the country’s largest cotton market. New York editorialists denounced abolitionism in the 1830s and blamed race riots on political agitation, which they called a threat to the city’s prosperity. On the eve of Southern secession, the southern writer J.D.B. DeBow wrote that New York was “almost as dependent upon Southern slavery as Charleston itself” (quoted in Quigley

2005, p. 283). Some recent historical writing accepts this rhetoric at face value: for example, “New Yorkers knew that they were dependent on southern cotton and the slave system that planted, picked and packaged it,” writes Wells (2020). Fernando Wood, who served three nonconsecutive terms as mayor of New York starting in 1854, apparently took these claims seriously as well, because he responded to the crisis of Southern secession by proposing that the New York secede from the Union as well, and set itself up as a free city.

In his book on New York City elites, however, Beckert (2014, pp. 90–91) wrote that even while merchants and bankers wanted to accommodate the South politically, a rising group of upper-class New Yorkers had come to believe that “the political power of southern slaveholders over the federal government was nothing less than a threat to the United States and to their own economic wellbeing ... Moreover, the political power of these southern slaveholders, these businessmen began to argue, prevented necessary reforms in the banking, currency, credit and transportation systems.” Such viewpoints in the seat of US financial capitalism signaled the growing community of economic interest between northeastern and northwestern regions, in which the slave South had no integral place, but instead stood in the way of progress.

The Republican Party emerged in the midwest in the 1850s with the primary goal of opposing extension of slavery into the territories. But the new party also advanced an activist, pro-growth agenda for the federal government, one that had been taking shape for decades. One part of the program, supported by northeastern manufacturers and opposed by the South, was the protective tariff. Another major item was an ambitious federal infrastructure plan, to foster agricultural development and improve access to markets. An important focal point was the Great Lakes transportation system, in many ways an extension of the Erie Canal project, but beyond the capacity of any one state. The South opposed such programs on the grounds that they would mainly benefit other regions. Southern and southern-aligned presidents vetoed seven “Rivers & Harbors” bills between 1840 and 1860 (Egnal 2009). Republican campaigns also featured support for agricultural research and education, measures long supported by a majority of northern farmers but opposed by the South (Ron 2016). Thus by the 1850s, the slave South stood in opposition to a long agenda: a Homestead Act; infrastructure investments including the Pacific Railroad; currency and banking reform; and federal support for agricultural research and education. Much of this program was in fact enacted with Lincoln’s election in 1860, and the departure of southern representatives from Congress.

Earlier “economic interpretations” of the Civil War, such as those advanced by Charles and Mary Beard (1927), made the mistake of minimizing the importance of slavery itself as a root cause of regional conflict: these arguments held that slavery was a hot-button moral debate, but the “real” issues were economic. To the contrary, real-world concerns reinforced the view that the slave South posed a positive threat to American economic progress. Northerners had long associated slavery with economic backwardness and with closing off opportunities for small farmers and settlers. “Free labor requires free soil” was a common slogan (Morrison 1997,

p. 110). For areas whose slavery status was up for determination, the threat posed by slave-owning institutions to public education for the free population was widely appreciated (Majewski 2016).

There were elements of racism in free-soil rhetoric, a feature that should be candidly acknowledged. The *Illinois State Journal* wrote in 1856: “We are for free Kansas, because we are for free white men” (quoted in Morrison 1997, p. 172). White racism was a pervasive feature of nineteenth-century America, so this is not a matter of ranking the regions on a metric of virtue. It is about understanding the role of slavery in the country’s economic growth.

Conclusion

Neither slavery nor cotton were instrumental in the American growth acceleration of the nineteenth century. By contrast, political independence from Great Britain was a critical background factor, by allowing domestic financial institutions and capital markets to emerge, and by liberating state and local entities to sponsor “internal improvements” and other forms of social capital formation. Political independence also allowed the northern US states to abolish slavery, essentially the first abolitions in modern history. Exclusion of slavery under the Northwest Ordinance of 1787, in areas where the institution of slavery had many supporters and might well have flourished, was vital for the country’s growth trajectory.

New world slavery should be understood in context, as a device for profiting from newly conquered territories. As such, the practice made many fortunes and served Great Britain well, by fostering the expansion of transatlantic trade during the eighteenth century. Mainland North America was relatively peripheral for this colonial empire. However, the mainland’s favorable demographic environment meant that slave-owners in the southern states were well-positioned to seize the opportunities arising from the rise of cotton in the nineteenth century. For them, slavery “worked” in two distinct senses: they were able to transport unfree labor into prime cotton-growing areas, generating a rapid expansion of production; and slave-owners as a class grew rich in the process, their wealth capitalized in the value of slave property.

The incentives associated with this property, however, led slaveholders to eschew or neglect activities that fostered growth. As owners of scarce, valuable labor, they approved the closing of the African slave trade and discouraged recruitment of free settlers or workers. Because the value of their human property was independent of local development, they did not form local and regional coalitions to promote transportation and towns, as occurred in the free states. For similar reasons, slave-owners saw little benefit to educating the free population of the South and were positively fearful at the prospect of educating slaves. These policies or non-policies were clearly unfavorable for long-run development. The adverse consequences were already visible before the US Civil War. The slave South did not offer attractive, growing markets for farm products, middle-class consumer goods, or new

technologies comparable to those emerging from the family farms and cities of the northern states.

The debates over the effects of slavery on nineteenth-century US economic growth have revived recently within the context of national discussions of reparations for descendants of enslaved Americans. The reading of American economic history presented here has no particular bearing on these demands. Whether the role of slavery in US nineteenth-century economic ascendancy was major, minor, or negative, the historical facts would remain that African Americans labored involuntarily and were denied equal rights long after Emancipation. From this perspective, complicity was just as culpable as direct engagement with slavery. The injustice of slavery was all too real, and post-Civil War Reconstruction efforts at a fresh economic and political start proved historically fleeting. Calls for greater attention to this history should be applauded and encouraged, even though the intellectual terrain will inevitably be controversial. In debates over racial justice, exaggerated or misleading claims for a central or decisive role for slavery in US economic development can only distract from the core mission. Historical responsibility is ultimately a national affair, not something that can or should be inferred from growth statistics.

■ *The author thanks editors Heidi Williams, Erik Hurst, Nina Pavcnik, and especially Timothy Taylor, for careful readings and thoughtful suggestions on an earlier draft. Paul Rhode also offered helpful advice. The list of those with whom I have discussed and debated these issues over the years is too long to include here, but I am grateful to all of these scholars and friends.*

References

- Abbott, Richard H.** 1991. *Cotton and Capital: Boston Businessmen and Antislavery Reform, 1854–1868*. Amherst, MA: University of Massachusetts Press.
- Beard, Charles A., and Mary Ritter Beard.** 1927. *The Rise of American Civilization*. New York: The Macmillan Company.
- Beckert, Sven.** 2014. *Empire of Cotton: A Global History*. New York: Vintage Books.
- Bleakley, C. Hoyt, and Paul Rhode.** 2021. “The Economic Effects of American Slavery, Redux: Tests at the Border.” Paper presented at the Economic History Association Annual Meeting, Tucson AZ, October 29–31.
- Broadberry, Stephen, and Bishnupriya Gupta.** 2009. “Lancashire, India, and Shifting Competitive Advantage in Cotton Textiles, 1700–1850: The Neglected Role of Factor Prices.” *Economic History Review* 62 (2): 279–305.
- Burnard, T.G.** 2001. “‘Prodigious Riches’: The Wealth of Jamaica before the American Revolution.” *Economic History Review* 54 (3): 506–24.
- Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright.** 2006. *Historical Statistics of the United States: Millennial Edition*. New York: Cambridge

- University Press.
- Chaplin, Joyce E.** 1991. "Creating a Cotton South in Georgia and South Carolina, 1760–1815." *Journal of Southern History* 57 (2): 171–200.
- Clarkson, Thomas.** 1846. "Letter from Thomas Clarkson." In *The Liberty Bell*, edited by Maria Weston Chapman, 58–64. Boston: Massachusetts Anti-slavery Fair.
- Darity, William A., Jr.** 1982. "A General Equilibrium Model of the Eighteenth-Century Atlantic Slave Trade: A Least-Likely Test for the Caribbean School." In *Research in Economic History*, Vol. 7, edited by Paul Uselding, 287–325. Greenwich, CT: JAI Press.
- Davis, David Brion.** 1966. *The Problem of Slavery in Western Culture*. Ithaca, NY: Cornell University Press.
- Davis, Ralph.** 1962. "English Foreign Trade, 1700–1774." *Economic History Review* 15 (2): 285–303.
- Davis, Ralph.** 1979. *The Industrial Revolution and British Overseas Trade*. Leicester, UK: Leicester University Press.
- Desmond, Matthew.** 2019. "In Order to Understand the Brutality of American Capitalism, You Have to Start on the Plantation." *New York Times*, August 14. <https://www.nytimes.com/interactive/2019/08/14/magazine/slavery-capitalism.html>.
- Domar, Evsey D.** 1970. "The Causes of Slavery or Serfdom: A Hypothesis." *Journal of Economic History* 30 (1): 18–32.
- Drescher, Seymour.** 1987. *Capitalism and Antislavery: British Mobilization in Comparative Perspective*. New York: Oxford University Press.
- Egnal, Marc.** 2009. *Clash of Extremes: The Economic Origins of the Civil War*. New York: Hill and Wang.
- Eltis, David, and David Richardson, eds.** 2008. *Extending the Frontier: Essays on the New Transatlantic Slave Trade Database*. New Haven, CT: Yale University Press.
- Engerman, Stanley L.** 2017. "Review of *The Business of Slavery and the Rise of American Capitalism, 1815–1860* by Calvin Schermerhorn and *The Half Has Never Been Told: Slavery and the Making of American Capitalism* by Edward E. Baptist." *Journal of Economic Literature* 55 (2): 637–43.
- Farrow, Anne, Joel Lang, and Jennifer Frank.** 2005. *Complicity: How the North Promoted, Prolonged, and Profited from Slavery*. New York: Ballantine Books.
- Findlay, Ronald.** 1990. *The "Triangular Trade" and the Atlantic Economy of the Eighteenth Century: A Simple General Equilibrium Model*. Essays in International Finance series, vol. 177. Princeton, NJ: International Finance Section, Princeton University.
- Fishlow, Albert.** 1965. *American Railroads and the Transformation of the Antebellum Economy*. Cambridge MA: Harvard University Press.
- Foote, Thelma Wills.** 2004. *Black and White Manhattan: The History of Racial Formation in Colonial New York City*. New York: Oxford University Press.
- Galenson, David.** 1981. *White Servitude in Colonial America: An Economic Analysis*. New York: Cambridge University Press.
- Gallay, Alan.** 2002. *The Indian Slave Trade: The Rise of the English Empire in the American South, 1670–1717*. New Haven, CT: Yale University Press.
- Gallay, Alan, ed.** 2009. *Indian Slavery in Colonial America*. Lincoln, NE: University of Nebraska Press.
- Gallman, Robert E.** 1970. "Self-Sufficiency in the Cotton Economy of the Antebellum South." In *The Structure of the Cotton Economy of the Antebellum South*, edited by William N. Parker, 5–23. Washington, DC: Agricultural History Society.
- González, Felipe, Guillermo Marshall, Suresh Naidu.** 2017. "Start-up Nation? Slave Wealth and Entrepreneurship in Civil War Maryland." *Journal of Economic History* 77 (2): 373–405.
- Greene, Lorenzo J.** 1942. *The Negro in Colonial New England, 1620–1776*. New York: Columbia University Press.
- Gross, Daniel P.** 2016. "The Ties That Bind: Railroad Gauge Standards, Collusion, and Internal Trade in the 19th Century U.S." Harvard Business School Working Paper 17-044.
- Guasco, Suzanne Cooper.** 2001. "'The Deadly Influence of Negro Capitalists': Southern Yeoman and the Resistance to the Expansion of Slavery in Illinois." *Civil War History* 47 (1): 7–29.
- Harley, C. Knick.** 1992. "International Competitiveness of the Antebellum American Cotton Textile Industry." *Journal of Economic History* 52 (3): 559–84.
- Harley, C. Knick.** 2001. "The Antebellum Tariff: Different Products or Competing Sources? A Comment on Irwin and Temin." *Journal of Economic History* 61 (3): 799–805.
- Heerman, M. Scott.** 2018. *The Alchemy of Slavery: Human Bondage and Emancipation in the Illinois Country, 1730–1865*. America in the Nineteenth Century series. Philadelphia: University of Pennsylvania Press.

- Herbst, Lawrence A.** 1978. *Interregional Commodity Trade from the North to the South and American Economic Development in the Antebellum Period*. New York: Arno Press.
- Hilt, Eric.** 2017. "Economic History, Historical Analysis, and the 'New History of Capitalism'." *Journal of Economic History* 77 (2): 511–36.
- Irwin, Douglas A., and Richard Sylla, eds.** 2011. *Founding Choices: American Economic Policy in the 1790s*. Chicago: University of Chicago Press.
- Irwin, James R.** 1988. "Exploring the Affinity of Wheat and Slavery in the Virginia Piedmont." *Explorations in Economic History* 25 (3): 295–322.
- Johnson, Arthur M., and Barry E. Supple.** 1967. *Boston Capitalists and Western Railroads*. Cambridge, MA: Harvard University Press.
- Kettel, Thomas Prentice.** 1965. *Southern Wealth and Northern Profits*. Tuscaloosa, AL: University of Alabama Press, 1860.
- Kilbourne, Richard Holcombe, Jr.** 1995. *Debt, Investment, Slaves: Credit Relations in East Feliciana Parish, Louisiana, 1825–1885*. Tuscaloosa, AL: University of Alabama Press.
- Kilbourne, Richard Holcombe, Jr.** 2006. *Slave Agriculture and Financial Markets in Antebellum America: The Bank of the United States in Mississippi, 1831–1852*. London: Pickering and Chatto.
- King, Alvy L.** 1970. *Louis T. Wigfall: Southern Fire-Eater*. Baton Rouge, LA: Louisiana State University Press.
- Klein, Rachel N.** 1990. *Unification of a Slave State: The Rise of the Planter Class in the South Carolina Backcountry, 1760–1808*. Chapel Hill: University of North Carolina Press.
- Lakwete, Angela** 2003. *Inventing the Cotton Gin: Machine and Myth in Antebellum America*. Johns Hopkins Studies in the History of Technology series. Baltimore, MD: Johns Hopkins University Press.
- Lindert, Peter H., and Jeffrey G. Williamson.** 2016. *Unequal Gains: American Growth and Inequality since 1700*. Princeton Economic History of the Western World series. Princeton, NJ: Princeton University Press.
- Lindstrom, Diana L.** 1970. "Southern Dependence upon Interregional Grain Supplies: A Review of the Trade Flows, 1840–1860." In *The Structure of the Cotton Economy in the Antebellum South*, edited by William N. Parker. Washington DC: Agricultural History Society.
- Lindstrom, Diane L.** 1978. *Economic Development in the Philadelphia Region, 1810–1850*. New York: Columbia University Press.
- Lozier, John William.** 1986. *Taunton and Mason: Cotton Machinery and Locomotive Manufacture in Taunton, Massachusetts, 1811–1861*. New York: Garland.
- Majewski, John.** 2016. "Why Did Northerners Oppose the Expansion of Slavery? Economic Development and Education in the Limestone South." In *Slavery's Capitalism: A New History of American Economic Development* edited by Sven Beckert and Seth Rockman 277–98. Philadelphia: University of Pennsylvania Press.
- Martin, Bonnie.** 2010. "Slavery's Invisible Engine: Mortgaging Human Property." *Journal of Southern History* 76 (4): 817–66.
- Martin, Bonnie.** 2016. "Neighbor-to-Neighbor Capitalism: Local Credit Networks and the Mortgaging of Slaves." In *Slavery's Capitalism: A New History of American Economic Development* edited by Sven Beckert and Seth Rockman 107–21. Philadelphia: University of Pennsylvania Press.
- Marx, Karl.** 1963. *The Poverty of Philosophy*. New York: International Publishers.
- Mason, Matthew E.** 2000. "Slavery Overshadowed: Congress Debates Prohibiting the Atlantic Slave Trade to the United States, 1806–1807." *Journal of the Early Republic* 20 (1): 59–81.
- McCusker, John J., and Russell R. Menard.** 1985. *The Economy of British America, 1607–1789*. Chapel Hill: University of North Carolina Press.
- Menard, Russell R.** 2001. *Migrants, Servants, and Slaves: Unfree Labor in Colonial British America*. Aldershot: Ashgate..
- Merritt, Keri Leigh.** 2017. *Masterless Men: Poor Whites and Slavery in the Antebellum South*. Cambridge: Cambridge University Press.
- Moore, Sean D.** 2019. *Slavery and the Making of Early American Libraries: British Literature, Political Thought, and the Transatlantic Book Trade, 1731–1814*. Oxford: Oxford University Press.
- Morrison, Michael A.** 1997. *Slavery and the American West: The Eclipse of Manifest Destiny and the Coming of the Civil War*. Chapel Hill: University of North Carolina Press.
- Myers, Margaret G.** 1931. *The New York City Money Market, Vol. 1, Origins and Development*. New York: Columbia University Press.
- New York Times Magazine.** 2019. "The 1619 Project." August 19, <https://www.nytimes.com/interactive/2019/08/14/magazine/1619-america-slavery.html>.

- Nieboer, Herman Jeremias.** 1900. *Slavery as an Industrial System: Ethnological Researches*. The Hague: Martinus Nijhoff.
- North, Douglass C.** 1966. *The Economic Growth of the United States, 1790–1860*. New York: W.W. Norton and Company.
- O'Connor, Thomas.** 1968. *Lords of the Loom: The Cotton Whigs and the Coming of the Civil War*. New York: Charles Scribner's Sons.
- Olmstead, Alan L.** 1972. "Investment Constraints and New York City Mutual Savings Bank Financing of Antebellum Development." *Journal of Economic History* 32 (4): 811–40.
- Olmstead, Alan L., and Paul W. Rhode.** 2008. "Biological Innovation and Productivity Growth in the Antebellum Cotton Economy." *Journal of Economic History* 68 (4): 1123–71.
- Olmstead, Alan L., and Paul W. Rhode.** 2011. "Productivity Growth and the Regional Dynamics of Antebellum Southern Development." In *Economic Evolution and Revolution in Historical Time*, edited by Paul W. Rhode, Joshua L. Rosenbloom, and David F. Weiman, 180–213. Stanford, CA: Stanford Economics and Finance.
- Olmstead, Alan L., and Paul W. Rhode.** 2018. "Cotton, Slavery, and the New History of Capitalism." *Explorations in Economic History* 67: 1–17.
- Parker, William N., and Robert E. Gallman.** 1992. "Replication data for: Southern Farms Study, 1860." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/ICPSR07419.v1>.
- Quarles, Benjamin.** 1976. "Antebellum Free Blacks and the 'Spirit of '76.'" *Journal of Negro History* 61 (3): 229–42.
- Quigley, David.** 2005. "Southern Slavery in a Free City: Economy, Politics and Culture. In *Slavery in New York*, edited by Ira Berlin and Leslie M. Harris, 263–88. New York: The New Press.
- Ransom, Roger L., and Richard Sutch.** 1988. "Capitalists without Capital: The Burden of Slavery and the Impact of Emancipation." *Agricultural History* 62 (3): 133–60.
- Reséndez, Andrés.** 2016. *The Other Slavery: The Uncovered Story of Indian Enslavement in America*. New York: Houghton Mifflin Harcourt.
- Richardson, David.** 1991. "Slavery, Trade, and Economic Growth in Eighteenth-Century New England." In *Slavery and the Rise of the Atlantic System*, edited by Barbara L. Solow, 237–64. Cambridge, UK: Cambridge University Press.
- Rhode, Paul W. and Richard Sutch.** 2006. "Estimates of National Product Before 1929." In *Historical Statistics of the United States: Millennial Edition*, edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, 12–20. New York: Cambridge University Press.
- Ron, Ariel.** 2016. "Summoning the State: Northern Farmers and the Transformation of American Politics in the Mid-nineteenth Century." *Journal of American History* 103 (2): 347–74.
- Rosenberg, Morton M., and Dennis V. McClurg.** 1968. *The Politics of Pro-Slavery Sentiment in Indiana, 1816–1861*. Ball State Monograph Number series. Muncie, IN: Ball State University.
- Rosenberg, Nathan.** 1963. "Technological Change in the Machine Tool Industry, 1840–1910." *Journal of Economic History* 23 (4): 414–43.
- Rosenbloom, Joshua L.** 2018. "The Colonial American Economy." Department of Economics, Iowa State University Working Paper18002.
- Rousseau, Peter L., and Richard Sylla.** 2005. "Emerging Financial Markets and Early US Growth," *Explorations in Economic History* 42 (1): 1–26.
- Saunt, Claudio.** 2019. "Financing Dispossession: Stocks, Bonds, and the Deportation of Native Peoples in the Antebellum United States." *Journal of American History* 106 (2): 315–37.
- Schermerhorn, Calvin.** 2015. *The Business of Slavery and the Rise of American Capitalism, 1815–1860*. New Haven, CT: Yale University Press.
- Schermerhorn, Calvin.** 2016. "The Coastwise Slave Trade and a Mercantile Community of Interest." In *Slavery's Capitalism: A New History of American Economic Development*, edited by Sven Beckert and Seth Rockman, 209–23. Early American Studies series. Philadelphia: University of Pennsylvania Press.
- Sokoloff, Kenneth L.** 1986. "Productivity Growth during Early Industrialization: Evidence from the American Northeast, 1820–1860." In *Long-Term Factors in American Economic Growth*, edited by Stanley L. Engerman and Robert E. Gallman, 679–736. Chicago: University of Chicago Press.
- Sylla, Richard.** 1998. "US Securities Markets and the Banking System, 1790–1840." *Federal Reserve Bank of St. Louis Review* 80 (3): 83–98.

- US Census Bureau.** 1918. *Negro Population, 1790–1915*. Washington, DC: U.S. Census Bureau.
- Wallis, John Joseph, and Barry R. Weingast.** 2018. “Equilibrium Federal Impotence: Why the States and Not the American National Government Financed Economic Development in the Antebellum Era.” *Journal of Public Finance and Public Choice* 33 (1): 19–44.
- Weiss, Thomas.** 1993. “Long-Term Changes in US Agricultural Output per Worker, 1800–1900.” *Economic History Review* 46 (2): 324–41.
- Wells, Jonathan Daniel.** 2020. *The Kidnapping Club: Wall Street, Slavery, and Resistance on the Eve of the Civil War*. New York: Bold Type Books.
- Wilder, Craig Steven.** 2013. *Ebony and Ivy: Race, Slavery, and the Troubled History of America’s Universities*. New York: Bloomsbury Publishing.
- Wilkins, Mira.** 1989. *The History of Foreign Investment in the United States to 1914*. Cambridge MA: Harvard University Press.
- Winthrop, Adam.** 1947. “Winthrop Papers, Vol. 5, (1645–1649).” Boston: Massachusetts Historical Society, 1628.
- Wood, Gordon S.** 1992. *The Radicalism of the American Revolution*. New York: Knopf.
- Wright, Gavin.** 1974. “Cotton Competition and the Post-Bellum Recovery of the American South.” *Journal of Economic History* 34 (3): 610–35.
- Wright, Gavin.** 1986. *Old South, New South: Revolutions in the Southern Economy since the Civil War*. New York: Basic Books.
- Wright, Gavin.** 2006. *Slavery and American Economic Development*. Baton Rouge, LA: Louisiana State University Press.
- Wright, Gavin.** 2020. “Slavery and Anglo-American Capitalism Revisited.” *Economic History Review* 73 (2): 353–83.

Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children

Anna Aizer, Hilary Hoynes, and Adriana Lleras-Muney

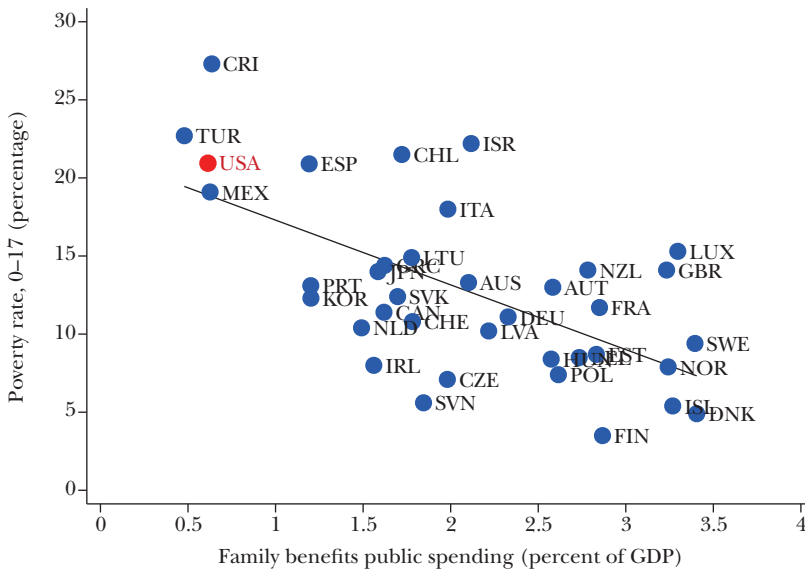
A hallmark of every developed nation is the provision of a social safety net—a collection of public programs that deliver aid to the poor. Because of their higher rates of poverty, children are often a major beneficiary of safety net programs. Countries vary considerably in both the amount of safety net aid to children and the design of their programs. The United States provides less aid to families with children as a share of GDP (0.6 percent) than most countries: Among 37 OECD countries, only Turkey provides less, as shown in Figure 1. Countries that provide less aid to families with children have higher rates of child poverty. Among these same 37 countries, only Turkey and Costa Rica have higher child poverty rates than the United States. Why does the United States appear to be such an outlier in terms of the amount of aid it provides to families and child poverty rates? While there are likely multiple reasons, in this paper we focus on one possible explanation: Past emphasis on the negative behavioral effects of safety net programs for families over the benefits of such programs for children.

The negative behavioral effects derive from the design of the safety net. The design of the social safety net is characterized by 1) who receives the aid, given

■ *Anna Aizer is Professor of Economics, Brown University, Providence, Rhode Island. Hilary Hoynes is Professor of Economics and Public Policy, Goldman School of Public Policy, University of California at Berkeley, Berkeley, California. Adriana Lleras-Muney is Professor of Economics, University of California Los Angeles, Los Angeles, California. All three authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are aizer@brown.edu, hoynes@berkeley.edu, and alleras@econ.ucla.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.149>.

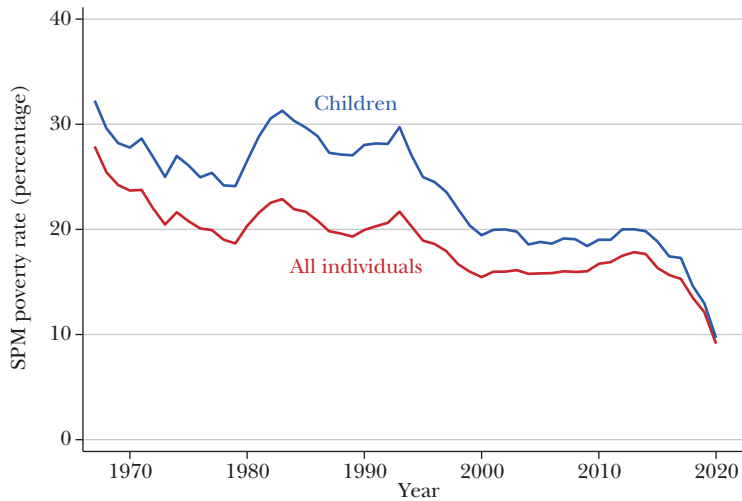
Figure 1

Child Poverty Rates and Public Spending on Families in OECD Countries

Source: Data come from the OECD database and use the latest information available mostly from 2017–2019. Poverty Rate Data (OECD 2022b) are available at <https://data.oecd.org/inequality/poverty-rate.htm>. Family Spending Data (OECD 2022a) are available at <https://data.oecd.org/socialexp/family-benefits-public-spending.htm>.

that some programs are universal while others are targeted to low-income families only; 2) how the aid is delivered, whether as cash or in-kind services such as housing subsidies or health insurance; and 3) whether the aid is conditional on some behavior such as work. When programs target low-income families, they deliver resources to those in greatest need. But this can potentially create negative incentive effects: for example, parents might change their behavior with respect to employment, marriage, or fertility to obtain or maintain eligibility. With respect to mode of delivery, economists tend to favor cash transfers because they are unconstrained and allow families to spend money to meet their greatest needs, but cash transfers can potentially result in spending on types of consumption that were not necessarily intended by the policymaker. This is particularly salient for programs that target children when parents are the recipients of the aid, because parents may not choose to spend the aid in ways that policymakers believe will improve the lives of their children, the intended target. Finally, safety net policies vary on whether aid is given based on certain conditions. In developing countries, conditional cash transfers often require parents to show that their children are attending school and/or receiving medical care. In the United States, the most common source of conditionality is work.

Figure 2

Trends in Poverty and Child Poverty in the United States, 1967–2020

Source: Unpublished data from Trisi and Saenz (2021) and using the methods described therein. The Supplemental Poverty Rate (SPM) is based on a family resource measure that includes cash income plus the value of in-kind transfers (food and nutrition and housing) minus taxes, selected work deductions, and out-of-pocket health payments. For detail, see Fox and Burns (2021).

Since 1967, rates of child poverty in the United States have fallen considerably when measured using the Supplemental Poverty Measure, shown in Figure 2. This measure counts the fraction of children in poverty after accounting for social safety net transfers (both in cash and in kind). Therefore, this measure allows one to evaluate the effectiveness of the safety net in removing children from poverty. Secular trends in child poverty observed in Figure 2 are strongly related to employment conditions (for example, they fell sharply in the strong economies of the 1990s and the 2010s), but are also related to changes in the size and structure of the social safety net.

Although the safety net for children in the United States has expanded over time, it has evolved away from unconditional cash transfers based on need to conditional transfers and to more in-kind benefits. A main objective of these changes has been to reduce negative behavioral responses on the part of the parents and to direct spending toward certain goods and services. Implicit or explicit work requirements, aimed to reduce the impact of the safety net programs on employment, have become increasingly common. However, this shift has the potential to reduce assistance to needy families. Indeed, many US families are either ineligible for safety net programs or fail to take up safety net programs for which they are eligible. Thus, while conditionality reduces disincentives and thus lowers costs, it also diminishes

the ability of the safety net to reduce poverty among children. Additionally, it is not clear whether conditioning aid on work (or other behaviors) is beneficial to the children of recipients.

Nevertheless, concern with behavioral effects has dominated not only discussions among policymakers but also economic research on the safety net for children. But the economic research on the safety net has evolved significantly over time, moving away from a near-exclusive focus on the negative incentive effects of the safety net on employment, earnings, marriage, and fertility to include examination of the potential positive benefits of such programs to children.

In this paper, we describe the evolution of this shift in economic research on the safety net to include benefits to children and the factors that precipitated the shift. We focus on the major safety net programs for children in the United States: cash welfare, food stamps, health insurance, and tax credits, listed in Table 1. The new research has shown that there are large benefits of safety net programs to children over the long run, with many programs proving to be excellent public investments that have paid for themselves (Hendren and Sprung-Keyser 2020). These findings should prompt policymakers to reassess the tradeoff between disincentive effects for adults and benefits for children.

Background on the Safety Net

Improving child health and wellbeing was the main objective of the first widespread US safety net program, the Mothers' Pension Program. In 1911, Illinois was the first state to enact such a program, with most states following soon thereafter, as illustrated by the timeline in Figure 3. These laws sought to provide mothers with dependent children (widows or those with incapacitated or incarcerated husbands) with an allowance that would enable children to remain at home and not be placed in institutional care or forced to work. Policymakers believed that this would, in turn, reduce child mortality and delinquency (Skocpol 1992).

During the Great Depression, states and counties were no longer able to fund their Mothers' Pension programs. In response, as part of the Social Security Act of 1935, the federal government created the Aid to Dependent Children (ADC) program. This program became the template for what is known as "welfare" in the United States. Originally, the ADC program (renamed Aid to Families with Dependent Children, or AFDC, in 1962) gave mothers cash on the basis of the number and ages of their children, as well as on the perceived needs of the family, which were assessed by social workers with substantial discretion. In most states, women were required to stay home (and not work) as a condition of the transfers. This approach was eventually replaced with a more formal system that guaranteed a minimum income, not a minimum transfer. As a result, the transfer was phased out at a very steep rate (referred to as the "benefit reduction rate") as beneficiaries earned income: The cash payments they received from the program decreased anywhere from 66 percent to 100 percent for every \$1 of earnings. In practice, benefits varied

Table 1
Major Safety Net Programs for Children in the United States

<i>Program</i>	<i>Description</i>	<i>Share of caseload that is children</i>	<i>Number of children served (2019, in millions)^a</i>	<i>Income eligibility</i>	<i>Estimated annual expenditures on children (2019, in billions)^b</i>
<i>Panel A. Major Programs (covered in this paper)</i>					
Medicaid/CHIP	Health insurance	0.51	35.0	<185% FPL for pregnant women and infants, <133% FPL for children (varies by state)	\$115.4
EITC	Tax credit for working families	0.74	33.0 (annual)	<150–235% FPL (depends on # of kids)	\$56.6
Child Tax Credit	Tax credit for families with children	1.00	—		\$117.7
SNAP	Food assistance through EBT card	0.44	19.9	<130% FPL	\$27.8
TANF	Block grant	0.75	1.5	<16%–150% FPL (varies by state)	\$12.8 ^c
AFDC (precursor to TANF, ended in 1995)	Unconditional cash transfer	0.68	9.1 (1995)	<185% of the state determined standard of need	\$40.0 (1996)
<i>Panel B. Other Programs that Serve Children</i>					
Supplemental Security Income	Cash transfer for aged and disabled	0.14	1.2	Countable income below \$750/month	\$10.5
Public housing ^d	Public housing and rental vouchers	0.45	4.5 (annual) ^e		\$8.5
Child Nutrition	School meals programs	1.00	29.6 ^f	≤130% FPL free and ≤185% FPL reduced price	\$22.3
Special Supplemental Food (WIC)	In-kind food assistance	0.76	4.8	<185% FPL for pregnant women and children <5	\$4.8

Note: See online Appendix for sources.

^aNumber of children served is average monthly in 2019 unless otherwise noted.

^bAnnual expenditures equal federal spending on children, based on estimates in Hahn et al (2021), unless noted otherwise.

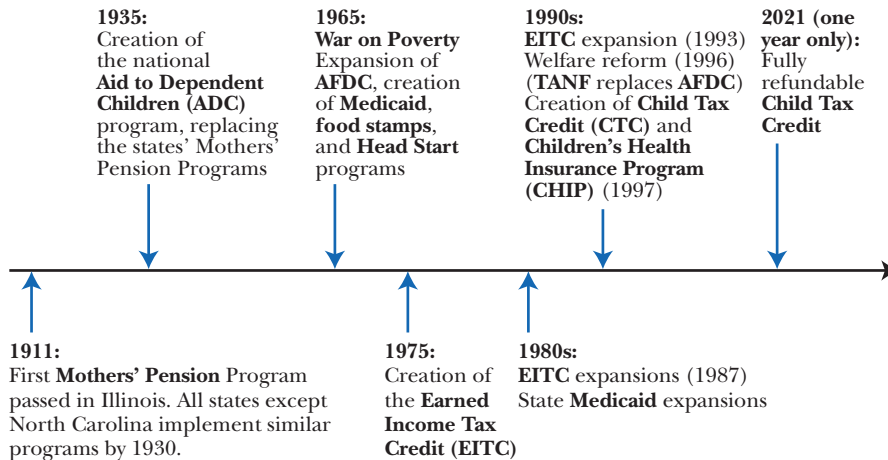
^cThis spending is for the entire block grant. An estimated 26 percent of the block grant is spent on cash assistance (Bitler and Hoynes 2016).

^dIncludes public housing and section 8 vouchers.

^eIncludes all households, not limited to households with children.

^fNumber of children served by school lunch program.

Figure 3

Evolution of Federal Anti-poverty Programs in the United States

Note: The Aid to Dependent Children Program's name was changed to Aid to Families with Dependent Children (AFDC) in 1962.

substantially by state and were quite modest; for example, in 1996 (the last year of the AFDC program before it was restructured and renamed as Temporary Assistance for Needy Families) *maximum* benefits averaged 35 percent of the federal poverty line across states (Safawi and Reyes 2021).

The safety net expanded substantially in the 1960s with President Lyndon Johnson's War on Poverty. Two additional in-kind programs to help the poor were established: food stamps and Medicaid (Bailey and Danziger 2013). Food stamps were provided to poor families to guarantee access to appropriate nutrition. Medicaid established free health insurance for poor families with children (and through later expansions, to the near-poor). There were also changes to the ADC program beginning as early as the 1950s and continuing into the 1960s and 1980s, with a goal of promoting work among mothers and lowering welfare dependency.

In the 1990s, two important policy changes pushed the safety net even further towards more work-based assistance (Hoynes and Schanzenbach 2018). The first was the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, commonly known as the welfare reform act, which replaced AFDC with Temporary Assistance for Needy Families (TANF) and significantly changed both the funding and requirements for the program. In the AFDC program, the federal government matched states' spending on cash assistance (at 50 percent or higher rates), whereas the new TANF program provided states with a block grant that was *fixed* in nominal terms and did not increase with enrollment (or prices). States have great flexibility in how to spend these funds. Finally, TANF imposed federal lifetime limits of 60 months for cash assistance and instituted work and training requirements for

recipients, though states can choose longer or, more commonly, shorter lifetime limits.

The second major change was the significant expansion of the Earned Income Tax Credit (EITC) in the early to mid-1990s and the creation of the Child Tax Credit in 1997. The EITC, first established in 1975, is a refundable tax credit for low-income working families with both eligibility and benefits increasing with the number of dependent children. Families with earned income pay lower taxes as a result and receive refund checks if their credit exceeds taxes owed. Families without earned income are not eligible. The program grew considerably in both eligibility and benefit levels starting in 1987 and accelerating during the 1990s. As a result of the expansion of the EITC and with the introduction of TANF, the cash-based social safety net became conditional on work. Like the EITC, the Child Tax Credit is a *conditional* tax credit received annually when individuals pay taxes. Individuals with little or no tax liability must have at least \$2,500 of annual earned income to qualify for the refundable portion of the credit.

Overall, the changes in cash assistance and the growth of the tax-based social safety net in the 1990s shifted focus from income support toward work-based assistance, and consequently the distribution of households receiving assistance moved away from the poorest households toward the near-poor (Hoynes and Schanzenbach 2018).

The Child Safety Net Today

Today, the largest safety net program for children is Medicaid, which currently provides health insurance to 35 million low-income children at an estimated cost of \$115 billion. (This total includes the Children's Health Insurance Program, introduced in 1996, which is targeted toward poor children.) Like AFDC and TANF, Medicaid is a state program that is regulated and partly funded at the federal level. Because the program covers health care costs, the actual value of the transfer to children varies greatly depending on the extent to which they use health care resources and how states determine eligibility and generosity.

The EITC is the second largest safety net program for children, providing an estimated \$56 billion in tax credits for families with children. It currently serves 33 million children in families with at least one working parent, transferring an annual credit worth \$3,204 on average. The Child Tax Credit is also large in terms of dollars spent, but most of the funds go to children that are not in poor families.

Next in size is the food stamp program—renamed as the Supplemental Nutrition Assistance Program (SNAP) program in 2008—which serves roughly 20 million children at a cost of \$30 billion. SNAP average monthly benefits are about \$130 per person.¹ The United States spends another \$45 billion on other in-kind programs for poor children, including housing, nutrition, and help for the disabled, as shown in Panel B of Table 1.

¹These figures are available from the US Department of Agriculture at <https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap>.

The traditional welfare program (TANF) is very modest today, currently serving 1.5 million children at a cost of about \$13 billion for the federal block grant. Both expenditures and caseloads for this program are lower today than in 1995 (as shown in Table 1). Moreover, there has been a movement away from cash assistance, with only 22 percent of TANF funds in 2020 being spent in this way, compared with 71 percent in 1997, pre-TANF (Bitler and Hoynes 2016; Safawi and Reyes 2021). The reductions in caseloads cannot be explained by reductions in the number of poor children: in 2019, for every 100 families with children in poverty, 23 participated in TANF, compared with 68 prior to TANF. The median monthly *maximum* benefit for a single-parent family of three is \$498 (about 30 percent of the poverty level). In practice, many poor families receive less than the maximum.

Evolution of Research on the Safety Net for Households with Children

The policy changes we have discussed were accompanied (or perhaps influenced) by important trends in the research on the impact of safety net programs. To characterize this evolution, we created a database of all articles employing empirical analyses of our core safety net programs (those listed in Panel A of Table 1) and published in the top general interest and field journals in economics since 1968. We identified 239 articles on the topic that we classify as estimating either negative incentive effects or benefits.² Of all the safety net programs, Medicaid received the most attention from economists, accounting for 41 percent of all articles, followed by traditional welfare programs (AFDC/TANF) at 34 percent, and SNAP at 18 percent. Overall, only 40 percent of papers estimated benefits from these programs.

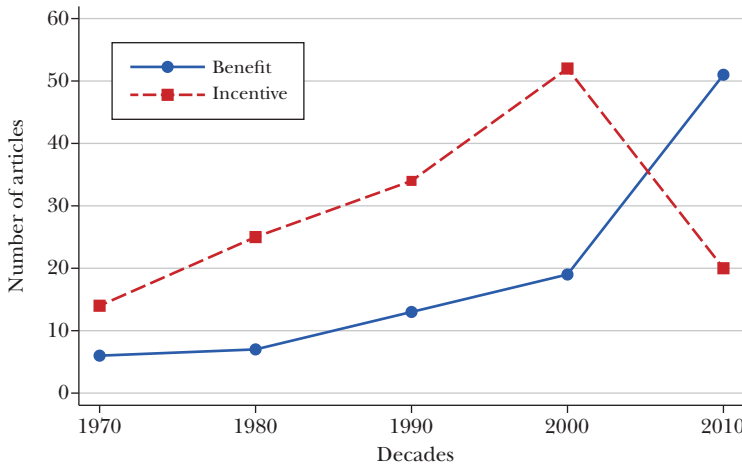
Figure 4 summarizes the number of papers published by decade and by objective. From the 1970s through the 2000s, the focus of economic research was the incentive effects of programs: prior to 2010, less than 27 percent of all articles documented benefits. Strikingly, in the last decade the research on benefits of safety net programs has taken off, with 2.5 articles on benefits for every article on incentives. This section describes the factors that led to this evolution.

Early Research on Safety Net Programs: Emphasis on Negative Incentive Effects

The optimal welfare program balances the benefits of increased consumption against the cost of the work disincentives it creates (Mirrlees 1971). For decades, in fact since the creation of the Mothers' Pension programs, concerns have been expressed by politicians and academics that welfare programs create disincentives for maintaining "traditional" family structures of marriage and work. Indeed, the theoretical predictions from standard economic models are unambiguous: The availability of a minimum "guarantee" of income and a benefit reduction rate is

²See notes to online Appendix Table 1 for details on journals included and selection criteria.

Figure 4

Research Articles on Incentive and Benefit Effects of Anti-Poverty Programs

Note: Data compiled by the authors. 2010 includes publications through 2020. See notes to Appendix Table 1 for details.

predicted to reduce employment and hours worked among female-headed households. Moreover, because the benefit increased with the number of dependent children in the household and declined with the presence of a married partner, this benefit schedule also created incentives to remain single and to have more children.

These concerns increased in the post-World War II years as a result of three broader demographic changes. First, the fraction of children growing up in single-parent households increased dramatically from 9 percent in 1960 to 20 percent in 1980 (US Census Bureau 2021). An increasingly large share of recipients were mothers who had never married (rather than divorced or widowed). Second, a larger share of recipients were Black, in part because many Black people migrated North, where welfare programs were more generous, and also because civil rights efforts were undertaken to increase access to welfare programs that had been systematically denied to Black people (Nadasen 2007). In contrast, prior to World War II, most recipients were white widows. In 1938, 48 percent of children on ADC were living with widowed mothers; by 1961 only 8 percent were. In 1938, 14 percent of the recipients were Black, but this share rose to 38 percent by 1956 and 44 percent by 1961 (Soule and Zylan 1997). Many observers believed that this growth in single-headed families was directly related to the incentives and generosity of AFDC in particular, a program whose expenditures and roles grew dramatically in the postwar years. Third, female labor-force participation among women with children rose steadily, in what Goldin (2006) labeled “the quiet revolution.” This shift raised the possibility that women with children could be self-sufficient and less dependent

on a safety net but were not doing so because of the incentives imbedded in the welfare programs.

One major effort to assess the validity of these concerns was the “income maintenance experiments,” four large-scale randomized experiments designed to assess the work disincentives of increasing the benefits and the benefit reduction rate of the AFDC program: in New Jersey (1968–72); North Carolina and Iowa (1969–73); Gary, Indiana (1971–74); and Seattle and Denver (1971–1982). These experiments, which economists were influential in designing and evaluating, randomized the income guarantee and the benefit reduction rate to a subset of families for three to ten years and tracked the resulting labor supply of the recipients. The experiments also included randomized subsidies for counseling and training. Although the results generally confirmed the predictions of economic theory that larger benefits and higher implicit tax rates lowered work, the estimated effects were surprisingly modest (for example, Robins 1985).

By 1992, the accumulated empirical evidence on the negative incentive effects of welfare was the subject of a review and synthesis in the *Journal of Economic Literature* (Moffitt 1992). Moffitt motivated the review with a reference in the introduction to the concerns of policymakers and the general public: “The US welfare system has been considered by many observers to be in a state of crisis since the late 1960s.” He cites two proximate causes of this crisis: the large growth in caseloads; and the implications of the growth “for possible long-term welfare dependency” and marriage disincentives that were seen as likely responsible for the growth in poor single-parent households over this period.

The available research at that time showed that the AFDC program generated “nontrivial” disincentives to work. For every \$1 in additional AFDC benefits, earnings were reduced by 37 cents. Though these disincentives seem strong, Moffitt (1992) concludes “the work disincentives of the program have little effect on the size of the caseload itself.” That is, even in the absence of the AFDC program, most women would have earnings that lie below the eligibility threshold. It is unclear why the labor supply responses to the negative work incentive in the welfare program are not as large as economists predicted. With respect to incentive effects regarding marriage and fertility, the results seem weak. As Moffitt (1992) summarizes: “The failure to find strong [cash] benefit effects is the most notable characteristic of this literature.”

Incorporating Benefits to Children in the Short Run

Absent from the Moffitt (1992) review is any reference to the potential *benefits* of cash assistance to the children of the recipients.³ The only reference to impacts on children in the review is to a small number of studies that estimate intergenerational welfare dependency. Indeed, the literature at the time included very few

³The same evolution of the research, with an initial focus on the effects on work, occurred for analysis of the Earned Income Tax Credit. Initial studies examined impacts on work and marriage, with little focus on effects on children (Hotz and Scholz 2003; Nichols and Rothstein 2016).

papers on the question, as Figure 4 shows. In our database, only two studies investigated effects of the influential income maintenance experiments on children in the 1970s (Maynard 1977; Maynard and Murnane 1979). Although these studies found that the experiment increased the education of children and young adults (at least in some samples), these findings were mostly ignored by policymakers and researchers alike. Perhaps economists ignored these because the popular Mirrlees (1971) framework is static, only considering the immediate consumption value of the transfers and ignoring any potential long-term benefits for children.

Starting in the mid-1990s, however, economists began to widen the focus of research on safety net programs to consider effects on household members, particularly children. One of the first papers in this new line of inquiry was “Welfare and Child Health: The Link between AFDC Participation and Birth Weight” (Currie and Cole 1993). As in Moffitt’s (1992) review, the authors motivate the paper with a reference to policy-making. They mention that many states had recently sought either to freeze or reduce payments in their AFDC programs, but were doing so “in a vacuum, because the effects of maternal participation on the wellbeing of their children has received little attention” (p. 971). Using a newly available data source containing information on welfare participation and child health—the National Longitudinal Survey of Youth panel—Currie and Cole (1993) estimated that AFDC benefits during pregnancy increased birth weights.

A second early example of research on the benefits of the social safety net to children comes from Medicaid. This too was motivated by policy debate: national health insurance reform was being debated in the United States in the mid-1990s. To shed light on the potential benefits of legislation that would expand health insurance universally, Currie and Gruber (1996a, b) studied the impact of the expansions in the Medicaid program in the 1980s which made more women and children eligible for the program. They found that increases in Medicaid coverage for pregnant women reduced infant mortality and the share of babies born of low birth weight (below 2,500 grams). These Medicaid expansions, which doubled the number of children eligible for Medicaid between 1984 and 1992, increased medical utilization and lowered child mortality. Since then, researchers have documented positive benefits of other safety net programs (the EITC and SNAP) during pregnancy on birth outcomes (for example, Strully, Rehkopf, and Xuan 2010; Hoynes, Miller, and Simon 2015; Almond, Hoynes, and Schanzenbach 2011).

Yet another set of early studies was made possible because of a series of randomized control trials of state welfare reforms in the early 1990s. Prior to the federal welfare reform that replaced AFDC with TANF in 1996, many states received waivers to reform their AFDC programs. Randomized control trials were used to examine impacts of these state reforms (Károly and Grogger 2005). While the first reports focused on effects on earnings and welfare participation, several papers extended that work to examine impacts of welfare reform on children, including health and school achievement (Gennetian et al. 2002; Morris, Gennetian, and Duncan 2005; Duncan, Morris, and Rodrigues 2011).

Linking Early Childhood with Long-Run Outcomes

Early work on how safety net programs like AFDC, Medicaid, food stamps, and the Earned Income Tax Credit affected child outcomes focused on birth outcomes, possibly because of data availability (birth weight and infant death are reported for all births in the vital registration systems) and also because it is easier to identify the effects of events that take place during the short gestation period. Advances in three areas of research helped to spur researchers' interest in examining the long-run effects of safety net use in childhood: research examining the long-run effects of preschool programs, examination of the long-run effects in utero and early-life circumstances, and work documenting low levels of economic mobility in the United States. We discuss each of these in turn.

Two influential and highly studied randomized control trials of preschool programs were conducted by psychologists in the 1960s and 1970s: the Perry Preschool Project and Abecedarian. These programs randomized children from disadvantaged households to high-quality preschool programs and followed them for 21 (Abecedarian) or 40 years (Perry Preschool). Both studies found substantial effects on academic achievement among children in the K–12 period, though these were not always statistically significant. Studies investigating participants as adults, however, showed a broad range of long-run benefits: Children randomized into the high-quality preschool program were more likely to complete high school, earn a bachelor's degree, and earn more in the labor market (Berruetta-Clement et al. 1984; Schweinhart and Weikart 1997a,b; Schweinhart et al. 2005; Heckman et al. 2010). Based on these long-run findings, cost-benefit analyses implied returns on investment on the order of 7–10 percent (Heckman et al. 2010). Importantly, many of the largest effects were not on educational outcomes, but on health and behavioral outcomes, with two-thirds of the financial returns coming from a reduction in criminal activity (Belfield et al. 2006). However, there were legitimate concerns over the generalizability of these findings. They had small sample sizes—around 100 participants in each study—all drawn from very poor families and with low cognitive test scores at baseline.

Following this work, several studies emerged evaluating the Head Start program, a program that provides free preschool for poor children, also pioneered by psychologists. Like the work estimating the impact of AFDC and Medicaid participation on birth outcomes, the work on Head Start also begins with a discussion of policymakers' desires to increase funding for Head Start even though “a careful reading of the literature reveals that credible studies demonstrating the lasting effects of Head Start are limited” (Currie and Thomas 1995, p. 341). The initial research documented short-term benefits as measured by improved test scores for children who participated compared to their siblings who did not (Currie and Thomas 1995). Interestingly, studies document significant fading of gains in the medium term (Puma et al. 2012), but significant long-term gains in educational attainment and earnings (Deming 2009; Garces, Thomas, and Currie 2002; Bailey et al. 2020).

A number of influential lessons emerged from this literature. First, there can be long-term positive benefits associated with high-quality preschool programs,

even if short- or medium-run effects are small or statistically insignificant. Second, we need to broaden our understanding of outcomes beyond a narrow focus on cognitive skills (like test scores) when considering child development, as there are large sizable returns to other forms of human capital. Third, accounting for both long-term effects and effects across a range of outcomes changes our evaluation of the desirability of these programs.

Similar lessons emerged from a different literature. In the early 2000s, economists began generating evidence linking in utero conditions, and birth weight, to long-term outcomes of social and economic significance. In particular, a seminal paper by Almond (2006) investigated whether children in the womb during the 1918 Spanish flu pandemic suffered negative consequences as a result of exposure to the virus in utero. Though epidemiologists had previously linked negative in utero environments and lower birth weight with worse *health* later in life (Barker et al. 1989), Almond (2006) showed that in utero flu exposure also lowered individuals' *economic well-being* as measured by educational attainment, disability, and earnings later in life. Many subsequent papers (reviewed in this journal by Almond and Currie 2011) have confirmed that exposure to the negative shocks in utero had negative long-term consequences for economic outcomes in adulthood.

A related line of research established a causal relationship between neonatal health and long-run outcomes using exogenous variation in birth weight *within pairs of twins* raised in the same household (Behrman and Rosenzweig 2004; Black, Devereux, and Salvanes 2007). The latter found that the effects of increased birth weight on newborn babies were much smaller than the long-term impacts on IQ, education, and earnings. This difference in findings (small short-run effects, but considerable long-run effects) further underscores the importance of examining multiple outcomes at different points in time to better understand how environments in early life influence child development.

This literature coincided with a third strand of research examining the relationship between paternal and child income. This work exploited new sources of data and documented much greater intergenerational correlations in earnings than had previously been estimated (Solon 1992): children of rich parents were much more likely to grow up to be rich than the children of poor parents. This research, since confirmed using new and better data, including that derived from IRS tax data (Chetty et al. 2014), implied lower rates of economic mobility in the United States, providing even more impetus for researchers to better understand how economic conditions in childhood shaped future outcomes.

Overall, this literature has led to a large number of papers investigating not only in utero shocks, but also how parental circumstances and policies in childhood affect economic and long-term well-being. In their extensive recent review, Almond, Currie, and Duque (2018) discuss the evidence on the long-term effects of childhood conditions, including the effects of policies and parental behaviors, not just external disease or environmental insults. This extended beyond programs like Head Start, explicitly targeting human capital investments, to include social

safety net programs such as food stamps, cash assistance, and the EITC, in which the primary recipient was an adult or family.

Recent Evidence on the Long-Run Benefits of Safety Net Programs for Children

New studies on the long-run impact of safety net programs have been made possible by advances in data collection and greater availability of data linking participation in safety net programs during childhood with long-term outcomes.

In an example of creatively accessing administrative data, Brown, Kowalski, and Lurie (2020) exploit tax data to estimate the long-term impact of Medicaid eligibility in childhood for children born 1981–1984 on a host of outcomes measured at ages 19–28. The authors find that those eligible for Medicaid in childhood enroll in college at higher rates, delay their fertility, experience reduced mortality, collect fewer EITC benefits, and pay higher taxes. Based only on the increased tax revenue, the authors estimate the government recoups roughly \$0.58 for every dollar spent on Medicaid eligibility in childhood. These findings were confirmed by Goodman-Bacon (2021), who finds that children who gained eligibility for Medicaid when the program was first introduced (1966–70) had better health and improved economic outcomes as adults. He concludes: “Childhood Medicaid coverage for these cohorts has therefore saved more than twice its cost.”⁴

Other work that uses existing data creatively to estimate long-term benefits of a safety net program on children is that of Hoynes, Schanzenbach, and Almond (2016) and Bailey et al. (2020). Hoynes, Schanzenbach, and Almond (2016) link existing data from the Panel Study of Income Dynamics on individuals born 1956–1981 with information on the staggered roll-out of the food stamp program across counties between 1961 and 1974. The authors estimate that gaining eligibility for food stamps during childhood improved long-term adult health and increased economic self-sufficiency among women. Bailey et al. (2020) use a similar identification strategy and take advantage of new linking technologies and data available at the US Census Bureau. They find that access to food stamps in early childhood leads to increases in completed education, earnings, neighborhood quality, and home ownership as well as reductions in poverty, mortality, and incarceration. In both these studies, the gains are large and increasing in length of exposure between conception and age five, after which there appear to be few effects, suggesting that early childhood may be a sensitive window for nutritional inputs.

Finally, Aizer et al. (2016) is an example of recent research that relies on the development of a new dataset to estimate the long-term impact of safety net programs. The authors collected and digitized data from historical archives that included information on all applicants to the Mothers’ Pension (1991–1935) program, and linked this information with mortality data, the 1940 Census, and World War II enlistment records for each boy in the sample. They estimate a positive

⁴Other recent work showing long term economic benefits of health insurance provision for children includes Cohodes et al. (2016), Miller and Wherry (2018), and Thompson (2017).

long-term impact of the original welfare program on children’s longevity, educational attainment, and earnings in young adulthood.

A number of other recent papers document these kinds of long-term benefits. How does this new information affect our evaluation of safety net programs? Recent work by Hendren and Sprung-Keyser (2020) systematically estimates the marginal value of public funds, taking into account the multiple long-term fiscal consequences of safety net programs. They document that programs targeted toward children have very large returns and essentially pay for themselves, in contrast to policies that target adults.

Implications for Policy-Making: Congressional Budget Office Scoring

How does research on the safety net translate into policy-making? Consider how the Congressional Budget Office scores policy proposals: Since 1975, every major piece of federal legislation has been evaluated by the CBO in an effort “to assess the effects on the economy of ‘major’ legislation that Congressional authorizing committees approve and to incorporate those effects into the agency’s *10-year cost estimates*” [emphasis added]. To do so, the CBO generates predictions for the overall cost of the legislation:

Cost estimates show how federal *outlays* and *revenues* would change if legislation was enacted and fully implemented as proposed—compared with what future spending and revenues would be under current law. Each estimate also includes a statement about the costs of any new federal mandates that the legislation would impose on state, local, or tribal governments or on the private sector.”⁵

This estimate—known as the CBO score—considers behavioral responses of individuals, firms, and local governments to the proposed legislation in calculating outlays and revenue.

A March 2015 report, “The Effects of Potential Cuts in SNAP Spending on Households with Different Amounts of Income,” illustrates the approach of the Congressional Budget Office to evaluating different policy options regarding a major safety net program, the Supplemental Nutrition Assistance Program—also known as “food stamps.” The main objective of this CBO (2015) analysis was to understand how three policies (reducing the maximum benefit by 13 percent, increasing the benefit reduction rate from 30 to 49 percent, and reducing the monthly income limit for eligibility from 130 to 67 percent of the federal poverty line), would affect outlays (costs) and household income, with the latter also potentially affecting tax revenue.

⁵ Accessed at CBO “Frequently Asked Questions About CBO Estimates” <https://www.cbo.gov/about/products/ce-faq>.

In the analysis, the CBO considered how a reduction in food stamp benefits would reduce direct outlays and affect parental labor supply. In predicting that labor supply would increase in response to benefit reductions, the CBO report cited Hoynes and Schanzenbach (2012), showing that after the food stamp program was introduced, female single heads of households worked fewer hours.

The Congressional Budget Office report did not, however, incorporate any predictions regarding changes in the *benefits* of participation associated with reducing SNAP participation. The CBO (2015) report concludes: “Participation in SNAP may have other consequences, such as effects on recipients’ health or nutrition, but evidence has so far been inconclusive” (p. 6). At the time of the report, however, evidence that participation in the food stamp program decreased the probability of low birth weight by between 5–11 percent did exist (Almond, Hoynes, and Schanzenbach 2011).

Why wasn’t this information included in the Congressional Budget Office report, given what we know about the short- and long-run benefits of reducing low birth weight on a number of important social and economic outcomes, including employment and earnings? A number of factors likely contribute to the omission. First, the CBO (as well as most economics researchers, historically) maintains a more narrow view of human capital that focuses on education and job training, areas in which the research linking skills with earnings is better developed. While the concept of “skills” with returns in the labor market expanded to include physical health, mental health, and other soft skills and behaviors in economics research, these concepts do not yet seem to have been fully incorporated into CBO estimation. One reason may be that measuring the exact rate of return to these other aspects of human capital is more complicated than measuring the private return to an additional year of schooling, which is now well established (Card 1999).

Second, the evidence presented in Almond, Hoynes, and Schanzenbach (2011), Hoynes, Schanzenbach, and Almond (2016), and Bailey et al. (2020) is based on historical data from the 1960s and 1970s. This generates concerns about generalizability to the present.

Third, and perhaps most importantly, the Congressional Budget Office is tasked with producing estimates *over a ten-year window*. Many of the returns to investments in children are not realized for many years, once the children complete their education, attain young adulthood, and enter the labor market. Thus, even if there were consensus on the long-run benefits of a program (which might need to be predicted if a program is new), the long-run benefits outside the ten-year window would not be included in the CBO scoring. In contrast, the costs of the program do register in the ten-year scoring window, as do any potential parental work disincentives. In sum, we capture the short-run costs but omit the long-term benefits to children because of the ten-year scoring window.⁶ This structure inhibits the ability of policymakers to take full advantage of recent advances in the economic research

⁶There are examples of programs that generated short-run effects that faded over the medium term, only to re-emerge when long-run outcomes were measured. This adds uncertainty to long-term projections.

exploring and documenting the long-run benefits of safety net programs, which often exceed their short-run costs by a wide margin (Hendren and Sprung-Keyser 2020).

A comparison of the estimated marginal value of public funds for the first welfare program in the United States, the Mothers' Pension program, in the short- and long-run provides an illustrative example of how time horizon influences this calculation. (The "marginal value of public funds" of a policy or program is calculated by estimating the benefits of the policy or program to recipients, measured as their willingness to pay, divided by the policy's net cost including long-term impacts on the government's budget.) The main estimated benefits of the Mothers' Pension program were increases in completed schooling and earnings in early adulthood as well as increases in longevity, which are not realized until many years after benefit receipt. In contrast, the main costs of the program are realized earlier in the form of direct outlays and delayed time to remarriage among the mothers. If one only considers the latter, the marginal value of public funds calculation is below 1 (at 0.84), suggesting that the costs of the program exceed its benefits. However, once benefits to children are considered, the marginal value of public funds rises to more than 5, suggesting that the program generates substantial benefits relative to the costs (Aizer, Eli, and Lleras-Muney 2020). The benefits are large, in part, because they accrue over many years and because there are substantial health benefits in addition to labor market benefits. Similarly, for the food stamp program, if one were only to consider the impact on adults, the marginal value of public funds would be 0.54. But once one considers evidence that food stamps increase the earnings and life expectancy and reduce incarceration of children, the marginal value of public funds including these impacts on children rises to 56 (Bailey et al. 2020).

In sum, these programs represent *investments* in the human capital of children, not simply transfers to adults that increase consumption. The returns of these investments, like those of other investments in human capital, can only be properly measured over the entire lifetime of the recipients and should be comprehensive in nature, including gains to schooling, health, and other aspects of human well-being.

How to Make Research More Informative for Policy-Making

There are two key areas in which future economic research can help improve the design and scope of the safety net for children in the United States: understanding why the estimated impacts of various programs on children seem to differ across settings; and shedding light on whether and how evidence based on short-term impacts can be used to predict long-term outcomes. We discuss each in turn.

Understanding Why Estimated Treatment Effects Differ across Settings

The effects of safety net programs on children's outcomes often vary based on population and environment. This makes it difficult for policymakers to generalize or extrapolate from one setting to another, and it impairs their ability to target

programs to those likely to benefit the most. For instance, we have highlighted work showing a positive impact of cash transfers through the safety net on child outcomes. But not all research finds positive effects of cash or in-kind transfers to the poor in the long run. For example, Price and Song (2018) find no long-term effects of the Seattle-Denver Income Maintenance Experiment on children's longevity or on economic outcomes in adulthood. But it is not clear why.

Documenting heterogeneity in the effects of the programs across settings should be accompanied by an exploration of its sources. A first likely source of heterogeneity is differences in the counterfactual environment faced by families eligible for safety net programs. For example, the benefits to children from participation in Head Start have been shown to depend on what the alternative source of care is, with the benefits of Head Start increasing for those whose alternative environment is more deprived or of lower quality (Kline and Walters 2016; Cascio 2021). Likewise, Medicaid expansions are more effective for the lowest-income mothers, which likely reflects the fact that higher-income mothers were more likely to have had earlier access to private insurance which Medicaid crowded out (Currie and Gruber 1996a).

A second potential source of heterogeneous effects is complementarities across multiple programs (program complementarity, the idea that one program is more effective if a child is also participating in another program), or across investments early and later in life (dynamic complementarity), for which the evidence is scant (Almond, Currie, and Duque 2018). Generating empirical evidence of complementarities across environments or programs is challenging due to the difficulty of obtaining plausibly exogenous variation in multiple inputs across time or programs. Johnson and Jackson (2019) make such an attempt and find that school finance reform is far more effective at raising student test scores if students had access to Head Start programming, and also that Head Start is more effective if followed by resource-rich educational programming. This might help explain why Black children have been found, in general, to benefit less in the long run than other groups from anti-poverty programs: If Black families are more likely to attend under-resourced or segregated schools, they may receive lower levels of complementary investments (for example, Garces, Thomas, and Currie 2002; Bailey et al. 2020).

Treatment effects may also vary because of interaction effects with labor market conditions or the housing market. For some children, this means that discrimination may also play an important role. Discrimination in the housing market, for example, could hinder the ability of Black and Latino families to leave high-poverty/high-crime neighborhoods with under-resourced schools. Discrimination in the labor market can reduce the returns to skills developed in childhood for Black and Latino youth. Discrimination may reduce access to safety net programs. A better understanding of observed differences by race and the role of discrimination is needed.

Recent work has attempted to better understand treatment effect heterogeneity by pooling results of similar experiments across settings in the spirit of meta-analysis (for example, Dehejia, Pop-Eleches, and Samii 2021; Meager 2019; Page 2021). These efforts highlight the difficulty of conducting such assessments.

Two barriers include non-standard measurement of outcomes across studies and lack of adequate information to allow researchers to pool the results across studies.⁷ Thus, standardized measurement of outcomes and providing more information about sample characteristics would help. Another strategy is to leverage newly available administrative data sets. Large samples allow researchers to estimate treatment effects more precisely for different subgroups. New machine learning methods designed to work with large data, such as those in Wager and Athey (2018), can be used to investigate heterogeneity systematically. Machine learning techniques allow researchers with large datasets to uncover new patterns of heterogeneity that are not obvious in advance in a statistically sound way.

Projecting Long-Term Treatment Effects from Short-Term Evidence

Some childhood shocks appear to have effects that “fade out” initially, only to re-appear later in life (Almond, Currie, and Duque 2018). An important next step for research is to improve our understanding of the extent to which evidence on short- and medium-term outcomes can be used to make long-term projections. This issue is particularly important given the short-term incentives inherent in policymaking. Programs that show no benefits in the short run tend to be eliminated, but these programs may ultimately “pay for themselves” if they generate long-term benefits. Conversely, programs with short-term benefits receive more political support, even if these benefits turn out to be short-lived. An exciting econometric development in this area is the use of new techniques that combine short-term evidence from randomized trials with long-term observational data to project long-run outcomes (Athey, Chetty, and Imbens 2020; Aizer, Eli, and Lleras-Muney 2021; García et al. 2020).

A related need is to develop a better understanding of what explains these dynamic effects. Estimated effects may “appear” at different points in time either because of dynamics in the underlying process of skill production (for example, a common pattern is that risky behavior does not manifest until adolescence and then declines in adulthood) or because data limitations prevent researchers from effectively measuring outcomes at different points in time. For example, measures of noncognitive skills are often crude, like measures of criminal involvement, or based on subjective measures, like answers on a questionnaire used to develop a behavioral problem index. As Guttmannova, Szanyi, and Cali (2008) discuss, these behavioral checklists often do poorly at characterizing the behavior of racial minorities and low-income children. To address this, economic research may benefit from insights from psychology or neurobiology regarding the timelines of child development. This can serve as a basis for collection of data on outcomes that better reflect the underlying developmental process.

⁷Card, Kluge, and Weber (2018) make similar observations about the difficulty of conducting such research in their paper, which combines the results of 207 papers to investigate the effects of work training programs. Galama, Lleras-Muney, and van Kippersluis (2018) report similar difficulties when attempting to explain the heterogeneity in the estimates of the causal effects of education on health.

A complementary effort is the development and empirical assessment of models that can generate the types of dynamic treatment effects that are observed in the data. The most important work in this area comes from the model of Cunha and Heckman (2007), which codified a number of existing insights and generated new ones about the formation of skill. Most notably their model allows for “sensitive windows,” during which certain skills are best produced, and for “dynamic complementarities,” where skills produced at an early stage raise the productivity of investment at subsequent stages. Another recent example comes from Lleras-Muney and Moreau (2020), who provide a model of health and mortality at the population level, where in utero shocks or differences in socioeconomic status throughout life can result in treatment effects that are U-shaped over time, as the literature has documented. This model accurately predicts the dynamic long-term effects of graduating in a recession on mortality (Schwandt and von Wachter 2019). Parallel models could be developed for other types of interventions and outcomes. Most importantly, empirical evidence in support of these models is needed.

Why Is the US Child Poverty Rate Still So High and What Can We Do about It?

The growth of the social safety net (particularly through tax credits and in-kind transfers) has played an important role in reducing child poverty but has not eliminated it, as shown earlier in Figure 2. In order to have large impacts on child poverty, a better understanding of the causes of poverty and how safety net programs influence poverty is needed.

There are multiple candidate explanations for the continued high rate of child poverty. The first and most obvious is that the United States spends less on assistance for families with children than other high-income countries, as discussed previously. The question is why. It’s not necessarily that the United States is unwilling to spend on social programs: The United States spends considerably more on the elderly than on children—7 percent of GDP for 54 million seniors compared with 0.5 percent of GDP for 73 million children—and indeed spends more on seniors (as a percent of GDP) than many other OECD countries. Moreover, the main programs for the elderly (Social Security retirement benefits and Medicare) are universal, while child benefits are income-targeted and include conditionality (such as work requirements). Indeed, initial estimates suggest that the United States reduced child poverty significantly in 2020 and 2021 through the expanded Child Tax Credit and other efforts related to pandemic relief, but the relief is temporary, with most of it expiring after 2021 (Urban Institute 2021; CCPS 2021).

Several factors may explain the disparity in US public funding between children and the elderly. The first is “senior power,” which refers to the political influence of the elderly. Seniors vote in large numbers—71 percent voter turnout in the 2016 election compared with 46 percent for ages 18 through 29—and are very well

organized. The American Association of Retired People (AARP) boasted 38 million members and \$1.7 billion in revenue in 2019. It constitutes a powerful lobby. On the other hand, children do not vote and there is no organization that represents the needs of children to rival the AARP. The Children's Defense Fund, one of the major groups advocating for children in the United States, reported revenue of \$17.8 million in 2019, just 1 percent of AARP revenue.

The second factor may be the racial and ethnic composition of the two populations. The elderly population in the United States is 77 percent white non-Hispanic, in contrast to children, who are slightly less than half white non-Hispanic. From the onset, the generosity and universality of anti-poverty programs have been a function of the racial composition of potential recipients. In the Mothers' Pension program, only 3 percent of all families receiving aid were Black, which was far less than what one would have predicted if aid were based solely on need (US Children's Bureau 1933). Many New Deal safety net programs of the 1930s explicitly excluded workers in industries with high Black employment shares (Katznelson 2005). This bias persisted throughout the 20th century. The 1996 welfare reform, which gave states considerable discretion over spending, led to lower levels of cash assistance in states with higher shares of Black residents (Hardy, Samudra, and Davis 2019). Today, states with large Black populations have substantially less generous welfare programs (Urban Institute 2021). This evidence is consistent with the conclusion of Alesina, Sacerdote, and Glaeser (2001) that race is the "single most important predictor of support for welfare" in the United States (see also Luttmer 2001). This combination of lack of political power and racial and ethnic discrimination have likely influenced the divergence in public priority and policy for these two groups.

A third reason is likely tied to the complexity of the safety net today. Children from low-income US households are often served by more than one program, and these programs are not well-coordinated. For example, 92 percent of children on food stamps participate in at least one other program and one-third participate in two other safety net programs (King and Giefer 2021). A call to coordinate and streamline eligibility for multiple safety net programs is not new (Currie 2006). A coordinated and comprehensive approach could work in a US context. It is notable that two of the most prominent and successful early education interventions (the Abecedarian and Perry School programs) provided a comprehensive set of services to children and their families.

Fourth, it may well be that even though the safety net has benefitted children, this complicated set of policies is not the best approach to lowering child poverty. Poverty rates for children are responsive to business cycles, with strong demand for labor and rising rates of employment and wages resulting in significant declines in child poverty. Perhaps labor market policies that increase the earned income of poor families, such as raising the minimum wage (Derenoncourt and Montialoux 2021), could prove to be as (or more) successful as direct aid. Another fruitful direction for research is to assess the relative cost-effectiveness of multiple approaches to lowering child poverty.

Conclusion

The emphasis of economic research on disincentives may have adversely influenced policy. Economic research on the effects of safety net programs has been and continues to be extremely influential in policy-making. For many years, this research focused primarily on documenting the potential negative behavioral impacts of these programs, even if small. Policymakers responded by attempting to minimize these behavioral impacts. The same focus on quantifying negative incentive effects has also historically dominated the study of social insurance programs, such as health insurance and unemployment insurance. The economic research on these programs has likewise evolved to include benefits as well as costs. A new focus on the benefits of safety net programs on children—the intended beneficiaries of these programs—that incorporates evidence on their many long-term benefits can contribute to the design of better policies going forward.

■ We thank Raheem Chaudhry, Sungwoo Cho, and Ray Huang, who provided excellent research assistance. We also thank Janet Currie, Greg Duncan, Shari Eli, Peter Germanis, and Robert Moffitt for helpful comments on an earlier draft.

References

- Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. “The Long-Run Impact of Cash Transfers to Poor Families.” *American Economic Review* 106 (4): 935–71.
- Aizer, Anna, Shari Eli, and Adriana Lleras-Muney. 2020. “The Incentive Effects of Cash Transfers to the Poor.” NBER Working paper 27523.
- Aizer, Anna, Shari Eli, Adriana Lleras-Muney, and Keyoung Lee. 2020. “Do Youth Employment Programs Work? Evidence from the New Deal.” NBER Working Paper 27103.
- Alesina, Alberto, Bruce Sacerdote, and Edward Glaeser. 2001. “Why Doesn’t the United States Have a European-Style Welfare State?” *Brookings Papers on Economic Activity* 32 (2): 187–278.
- Almond, Douglas. 2006. “Is the 1918 Influenza Pandemic Over? Long-Term Effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population.” *Journal of Political Economy* 114 (4): 672–712.
- Almond, Douglas, and Janet Currie. 2011. “Killing Me Softly: The Fetal Origins Hypothesis.” *Journal of Economic Perspectives* 25 (3): 153–72.
- Almond, Douglas, Janet Currie, and Valentina Duque. 2018. “Childhood Circumstances and Adult Outcomes: Act II.” *Journal of Economic Literature* 56 (4): 1360–446.
- Almond, Douglas, Hilary Hoynes, and Diane Schanzenbach. 2011. “Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes.” *Review of Economics and Statistics*, 93 (2): 387–403.
- Athey, Susan, Raj Chetty, and Guido Imbens. 2020. “Combining Experimental and Observational Data to Estimate Treatment Effects on Long Term Outcomes.” Unpublished.
- Bailey, Martha J., and Sheldon Danziger, eds. 2013. *Legacies of the War on Poverty*. National Poverty Center Series on Poverty and Public Policy. New York: Russell Sage Foundation.

- Bailey, M. J., Hilary W. Hoynes, Maya Rossin-Slater, and Reed Walker. 2020. "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program." NBER Working Paper 26942.
- Barker, D. J., P. D. Winter, C. Osmond, B. Margetts, and S. J. Simmonds. 1989. "Weight in Infancy and Death From Ischaemic Heart Disease." *Lancet* 334 (8663): 577–80.
- Behrman, Jere R., and Mark R. Rosenzweig. 2004. "Returns to Birthweight." *Review of Economics and Statistics* 86 (2): 586–601.
- Belfield, C. R., Nores, M., Barnett, S., and Schweinhart, L. J. 2006. "The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup." *Journal of Human Resources* 41 (1): 162–90.
- Berrueta-Clement, J. R., L.J. Schweinhart, W. S. Barnett, A. E. Epstein, and D. P. Weikart. 1984. *Changed Lives: The Effects of the Perry Preschool Program on Youths through Age 19*. Monographs of the High/Scope Educational Research Foundation. Ypsilanti: High/Scope Press.
- Bitler, Marianne, and Hilary Hoynes. 2016. *Strengthening Temporary Assistance for Needy Families*. Washington, DC: The Hamilton Project.
- Black, Sandra, Paul Devereux, and Kjell G. Salvanes. 2007. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics*, 122 (1): 409–39.
- Brown, David W., Amanda E. Kowalski, and Ithai Z. Lurie. 2020. "Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood." *Review of Economic Studies* 87 (2): 792–821.
- Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, Vol. 3, Part A, edited by Orley C. Ashenfelter and David Card, 1801–63. Amsterdam, Netherlands: Elsevier Science.
- Card, David, Jochen Kluge, and Andrea Weber. 2018. "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." *Journal of the European Economic Association* 16 (3): 894–931.
- Cascio, Elizabeth U. 2021. "Early Childhood Education in the United States: What, When, Where, Who, How, and Why." NBER Working Paper 28722.
- Columbia Center on Poverty and Social Policy (CCPS). 2021. *Research Roundup of the Expanded Child Tax Credit: The First 6 Months*. New York: Columbia Center on Poverty and Social Policy.
- Center on Budget and Policy Priorities (CBPP). 2021. *Policy Basics: Temporary Assistance for Needy Families*. Washington, DC: Center on Budget and Policy Priorities.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review* 104 (5): 141–47.
- Cohodes, Sarah R., Daniel S. Grossman, Samuel A. Kleiner, and Michael F. Lovenheim. 2016. "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions." *Journal of Human Resources* 51 (3): 727–59.
- Congressional Budget Office (CBO). 2015. *The Effects of Potential Cuts in SNAP Spending on Households with Different Amounts of Income*. Washington, DC: Congressional Budget Office.
- Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.
- Currie, Janet. 2006. *The Invisible Safety Net: Protecting the Nation's Poor Children and Families*. Princeton: Princeton University Press.
- Currie, Janet, and Nancy Cole. 1993. "Welfare and Child Health: The Link between AFDC Participation and Birth Weight." *American Economic Review* 83 (4): 971–85.
- Currie, Janet, and Jonathan Gruber. 1996. "Health Insurance Eligibility, Utilization of Medical Care, and Child Health." *Quarterly Journal of Economics* 111 (2): 431–66.
- Currie, Janet, and Jonathan Gruber. 1996. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy* 104 (6): 1263–96.
- Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85 (3): 341–64.
- Dehejia, Rajeev, Christian Pop-Eleches, and Cyrus Samii. 2021. "From Local to Global: External Validity in a Fertility Natural Experiment." *Journal of Business & Economic Statistics* 39 (1): 217–43.
- Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1 (3): 111–34.
- Derenoncourt, Ellora, and Claire Montialoux. 2021. "Minimum Wages and Racial Inequality." *Quarterly Journal of Economics* 136 (1): 169–228.

- Doyle, Joseph J., Jr.** 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review* 97 (5): 1583–610.
- Duncan, Greg J., Pamela A. Morris, and Chris Rodrigues.** 2011. "Does Money Really Matter? Estimating Impacts of Family Income on Young Children's Achievement with Data from Random-Assignment Experiments." *Developmental Psychology* 47 (5), 1263–79.
- Fox, Liana, and Kalee Burns.** 2021. *The Supplemental Poverty Measure: 2020*. Washington, DC: United States Census Bureau.
- Galama, Titus, Adriana Lleras-Muney, and Hans van Kippersluis.** 2018. "The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence." In *Oxford Research Encyclopedia of Economics and Finance*.
- Garces, Eliana, Duncan Thomas, and Janet Currie.** 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92 (4): 999–1012.
- Garcia, Jorge Luis, James Heckman, Duncan Ermini Leaf, and María José Prados.** 2020. "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program." *Journal of Political Economy* 128 (7): 2502–41.
- Garfinkel, Irwin, Laurel Sariscsany, Elizabeth Ananat, Sophie Collyer, and Christopher Wimer.** 2021. "The Costs and Benefits of a Child Allowance." Center on Poverty and Social Policy, February 18. www.povertycenter.columbia.edu/publication/2021/child-allowance/cost-benefit-analysis-discussion.
- Gennetian, Lisa A., Greg J. Duncan, Virginia W. Knox, Wanda G. Vargas, Elizabeth Clark-Kauffman, and Andrew S. London.** 2002. *How Welfare and Work Policies for Parents Affect Adolescents: A Synthesis of Research*. New York: Manpower Demonstration Research Corporation.
- Goldin, Claudia.** 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review* 96 (2) 1–21.
- Goodman-Bacon, Andrew.** 2021. "The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes." *American Economic Review* 111 (8): 2550–93.
- Gruber, Jonathan.** 2003. "Medicaid." In *Means-Tested Transfer Programs in the United States*, edited by Robert A. Moffitt, 15–77. Chicago: University of Chicago Press.
- Guttmanova, Katarina, Jason M. Szanyi, and Philip W. Cali.** 2008. "Internalizing and Externalizing Behavior Problem Scores: Cross-Ethnic and Longitudinal Measurement Invariance of the Behavior Problem Index." *Educational and Psychological Measurement* 68 (4): 676–94.
- 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.
- Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz.** 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94 (1–2): 114–28.
- Hendren, Nathaniel, and Ben Sprung-Keyser.** 2020. "A Unified Welfare Analysis of Government Policies." *Quarterly Journal of Economics* 135 (3): 1209–318.
- Hotz, V. Joseph, and John Karl Scholz.** 2003. "The Earned Income Tax Credit." In *Means-Tested Transfer Programs in the United States*, edited by Robert A. Moffitt. Chicago: University of Chicago Press.
- Hoynes, Hilary, Doug Miller, and David Simon.** 2015. "Income, the Earned Income Tax Credit, and Infant Health." *American Economic Journal: Economic Policy* 7 (1): 172–211.
- Hoynes, H., and D. W. Schanzenbach.** 2018. "Safety Net Investments in Children." *Brookings Papers on Economic Activity*, 89–150.
- 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.
- Hoynes, Hilary W., and Douglas 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 W., and Diane Whitmore Schanzenbach.** 2012. "Work Incentives and the Food Stamp Program." *Journal of Public Economics* 96 (1–2): 151–62.
- Johnson, Rucker C., and C. Kirabo Jackson.** 2019. "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending." *American Economic Journal: Economic Policy* 11 (4): 310–49.
- Karoly, Lynn A., and Jeffrey Grogger.** 2005. *Welfare Reform: Effects of a Decade of Change*. Cambridge, MA: Harvard University Press.

- Katznelson, Ira.** 2005. *When Affirmative Action Was White: An Untold History of Racial Inequality in Twentieth-Century America*. New York: Norton.
- King, Michael D., and Katherine Giefer.** 2021. *Nearly a Third of Children Who Receive SNAP Participate in Two or More Additional Programs*. Washington, DC: United States Census Bureau.
- Kline, Patrick, and Christopher R. Walters.** 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics* 131 (4): 1795–848.
- Lleras-Muney, Adriana, and Flavien E. Moreau.** 2020. "A Unified Model of Cohort Mortality for Economic Analysis." NBER Working Paper 28193.
- Luttmer, Erzo F.** 2001. "Group Loyalty and the Taste for Redistribution." *Journal of Political Economy* 109 (3): 500–28.
- Mallar, C.** 1977. "The Educational and Labor Supply Responses of Young Adults on the Urban Graduated Work Incentive Experiment." In *The New Jersey Income Maintenance Experiment: Expenditures, Health, and Social Behavior; and the Quality of the Evidence*, Vol. III, edited by H. W. Watts and A. Rees, 163–184. New York: Academic Press.
- Maynard, Rebecca A.** 1977. "Some Aspects of Income Distribution: The Effects of the Rural Income Maintenance Experiment on the School Performance of Children." *American Economic Review* 67 (1): 370–75.
- Maynard, Rebecca A., and Richard J. Murnane.** 1979 "The Effects of a Negative Income Tax on School Performance: Results of an Experiment." *Journal of Human Resources* 14 (4): 463–76.
- Meager, Rachael.** 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics* 11 (1): 57–91.
- Miller, Sarah, and Laura R. Wherry.** 2018. "The Long-Term Effects of Early Life Medicaid Coverage." *Journal of Human Resources* 54 (3): 785–824.
- Mirrlees, J. A.** 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.
- Moffitt, Robert.** 1992. "Incentive Effects of the U.S. Welfare System: A Review." *Journal of Economic Literature*, 30 (1): 1–61.
- Morris, Pamela A., Lisa A. Gennetian, and Greg J. Duncan.** 2005. "Effects of Welfare and Employment Policies on Young Children: New Findings on Policy Experiments Conducted in the Early 1990s." *Social Policy Report* 19 (2): 3–18.
- Nadasen, Premilla.** 2007. "From Widow to 'Welfare Queen': Welfare and the Politics of Race." *Black Women, Gender + Families* 1 (2): 52–77.
- 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. Chicago: University of Chicago Press.
- OECD.** 2022a. "Family Benefits Public Spending (indicator)." OECD Data. doi: 10.1787/8e8b3273-en (accessed March 29, 2022).
- OECD.** 2022b. "Poverty Rate (indicator)." OECD Data. doi: 10.1787/0fe1315d-en (Accessed on March 29, 2022).
- Page, M.** 2021. "What Have Economists Learned about Whether Money Matters to Child Well-Being?" Unpublished.
- Price, David J., and Jae Song.** 2018. "The Long-Term Effects of Cash Assistance." Princeton University Working Paper 621.
- Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer.** 2012. *Third Grade Follow-Up to the Head Start Impact Study*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Robins, Philip.** 1985. "A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments." *Journal of Human Resources* 20 (4): 567–82.
- Safawi, Ali, and Cindy Reyes.** 2021. *States Must Continue Recent Momentum to Further Improve TANF Benefit Levels*. Washington, DC: Center on Budget and Policy Priorities.
- Schwandt, Hannes, and Till von Wachter.** 2019. "Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets." *Journal of Labor Economics* 37 (S1): S161-S198.
- Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and**

- Milagros Nores.** 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Monographs of the High/Scope Educational Research Foundation. Ypsilanti: High/Scope Press.
- Schweinhart, Lawrence J., and David P. Weikart.** 1997. "The High/Scope Preschool Curriculum Comparison Study Through Age 23." *Early Childhood Research Quarterly* 12: 117–43.
- Schweinhart, Lawrence J., and David P. Weikart.** 1997. *Lasting differences: The High/Scope Preschool Curriculum Comparison Study Through Age 23*. Monographs of the High/Scope Educational Research Foundation. Ypsilanti: High/Scope Foundation.
- Skocpol, Theda.** 1992. *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States*. Cambridge, MA: Harvard University Press.
- Solon, Gary.** 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82 (3): 393–408.
- Soule, Sarah A., and Yvonne Zylan.** 1997. "Runaway Train? The Diffusion of State-Level Reform in ADC/AFDC Eligibility Requirements, 1940–1967." *American Journal of Sociology* 103 (3).
- Strully, Kate W., David H. Rehkopf, and Ziming Xuan.** 2010. "Effects of Prenatal Poverty on Infant Health: State Earned Income Tax Credits and Birth Weight." *American Sociological Review* 75 (4): 534–62.
- Trisi, Danilo, and Matt Saenz.** 2021. *Economic Security Programs Reduce Overall Poverty, Racial and Ethnic Inequities*. Washington, DC: Center on Budget and Policy Priorities.
- Urban Institute.** 2021. *2021 Poverty Projections: Assessing the Impact of Benefits and Stimulus Measures*. Washington, DC: Urban Institute.
- US Children's Bureau.** 1933. *Mother's Aid, 1931*. Bureau Publication 220. Washington: US Govt. Print. Off.
- US Census Bureau.** 2021. "Historical Living Arrangements of Children." Current Population Survey. <https://www.census.gov/data/tables/time-series/demo/families/children.html> (accessed on January 1, 2022).
- Wager, Stefan, and Susan Athey.** 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42.

Universal Early-Life Health Policies in the Nordic Countries

Miriam Wüst

Policymakers in the United States frequently look to the Nordic countries as a point of comparison. One area of interest is the well-documented infant health disadvantage in the United States relative to the Nordics: Despite high levels of health care spending and high costs of birth-related health care in the United States, an additional two to four infants per 1,000 die during their first year of life in the United States relative to the Nordic countries. As documented in Table 1, this infant mortality disadvantage is not exclusively driven by mortality in the narrow 28-day window after birth (the neonatal period) and persists during early childhood. Recent research has hypothesized that the absence of low-complexity preventive care during early life and offered at scale, like nurse home visiting for all new families, may be one factor explaining this infant health disadvantage (Chen, Oster, and Williams 2016).

While there are many other dissimilarities, stark differences in the maternal and child health care systems across the United States and the Nordic countries exist. For example, in the public health care systems of the Nordic countries—Denmark, Norway, Sweden, Iceland, and Finland—all families have access to a comprehensive set of standard programs free of charge at the time of service: prenatal care, care around time of birth, and well-child care in the first years of life. In contrast to the Nordic countries, no large-scale publicly funded and universally accessible early-life health programs are in place in the United States. Instead, existing US

■ *Miriam Wüst is Associate Professor of Economics, University of Copenhagen, Copenhagen, Denmark and CEBI-The Center for Economic Behavior and Inequality. Her email address is miriam.w@econ.ku.dk.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.175>.

Table 1

Some Economic and Mortality Indicators for the United States and the Nordic Countries

	<i>United States</i>	<i>Denmark</i>	<i>Finland</i>	<i>Iceland</i>	<i>Norway</i>	<i>Sweden</i>
GDP per capita, current prices, US dollars (2019)	\$65,056	\$58,738	\$50,376	\$58,347	\$66,870	\$54,657
Total health expenditure, share of GDP (2019)	16.8%	10%	9.2%	8.6%	10.5%	10.9%
Maternal mortality, % (2016)	0.02	<0.01	<0.01	<0.01	<0.01	<0.01
Neonatal mortality, % (2019)	0.35	0.26	0.14	0.1	0.14	0.15
Infant mortality, % (2019)	0.55	0.32	0.19	0.15	0.19	0.22
Child mortality, % (2019)	0.65	0.38	0.23	0.2	0.23	0.27

Source: GDP per capita and health care expenditure data from OECD.stat at stats.oecd.org. Maternal mortality is from Our World in Data at <https://ourworldindata.org/grapher/maternal-mortality-vs-neonatal-mortality?zoomToSelection=true&time=2016&country=DNK~SWE~NOR~ISL~FIN~USA>. Neonatal, infant, and child mortality are from UN Inter-agency Group for Child Mortality Estimation at <https://childmortality.org/>.

public policies focus on targeted programs, such as access to care covered under public health insurance for eligible pregnant women, infants, and small children or nutrition support in the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) program. Even near-universally accessible types of care, like prenatal care for pregnant women and hospital care at birth, are available to US families on different terms depending on factors such as insurance coverage. Finally, availability and design of US programs vary considerably across states and communities (such as a range of state and local home visiting programs for disadvantaged families). These differences across the United States and the Nordics may reflect both less interventionist political traditions and the absence of a universal public health insurance in the United States to fund universal care (Kamerman and Kahn 1993).

This essay asks what we know from economics research about the causal effects of Nordic universal early-life health policies. The aims of these policies as provided for general populations of families and their infants are to monitor health, to screen for maternal and infant health issues (and refer families with detected risks to other treatments), and to provide health information and counseling for all new parents to support parental investments. The essay starts by describing central features of the Nordic model and then summarizes lessons from research on universal prenatal care, care around time of birth, and well-child care for new families. The focus is on Denmark, Sweden, and Norway, where the existing work within economics predominantly is set. Given its focus on universal policies in the Nordic countries, the paper complements earlier reviews on targeted programs in the United States (Gomby, Culross, and Behrman 1999; Currie and Rossin-Slater 2015).

One type of evidence focuses on the extensive margin—the access to any care. The introduction of universal prenatal, maternity, and well-infant care preceded

the major expansion of the encompassing Nordic welfare states in the 1950s and 1960s and constitutes some of the first universal welfare policies still in place today. As such, evidence from policies introduced in this setting—with limited access to other health services and health insurance—is relevant for similar contexts today. It documents that the introduction of universal maternity and well-infant care in the Nordics causally contributed to the large decrease in infant mortality observed in the first half of the twentieth century and had important long-run benefits along dimensions of adult mortality, morbidity, education, and labor market outcomes. Findings from this research constitute one important part of a large literature documenting short- and long-run returns to large-scale programs that invest in the health and development of people early in life.

Another category of evidence comes from studies examining changes within universal programs today. This evidence is likely to be relevant for policymakers in many developed countries, where decisions revolve around the design of care rather than access to care. This research is still scarce and for now documents some positive impacts of universal early-life care on primarily short- and medium-run child and maternal health and a narrow set of parental investments (like breastfeeding or vaccination uptake). Universal screenings and provision of information and counseling for new parents appear to be important mechanisms. A general takeaway from studies on the Nordic universal health programs is that we should factor in returns over the life cycle, at non-mortality margins, and for both children and parents when assessing the cost effectiveness of universal early-life health policies (Evans et al. 2021).¹

The Nordic Model of Universal Early-Life Health Policies

A large empirical literature in economics documents the importance of negative health shocks during the in utero period and the first years of life for long-run health and economic outcomes. A nontrivial part of this work comes from studies set in the Nordic countries. For example, Black, Devereux, and Salvanes (2007) study variation in birthweight across identical twins in Norway; Nilsson (2017) examines prenatal exposure to strong beer in certain areas of Sweden in the 1960s; Persson and Rossin-Slater (2018) consider the effects of prenatal exposure to the death of a maternal relative in Swedish data; Black et al. (2019) focus on variation in radioactive exposure of children in Norway, due to nuclear weapons testing, during the 1950s and early 1960s; and Grönqvist, Nilsson, and Robling (2020) scrutinize the differential exposure of Swedish children to the phase-out of leaded gasoline. An

¹The Nordic welfare states today have a number of other features directed at families with small children, including maternity/parental leave programs and universal childcare. As a starting point for research on leave programs and childcare in the Nordic countries, the interested reader might begin with Rasmussen (2010), Carneiro, Løken, and Salvanes (2015), Havnes and Mogstad (2011, 2015), Datta Gupta and Simonsen (2010, 2016), and Dietrichson, Kristiansen, and Viinholt (2020).

Table 2

Main Features of Universal Early-Life Health Policies in Denmark, Norway, and Sweden in 2020

	<i>Denmark</i>	<i>Norway</i>	<i>Sweden</i>
Prenatal care	Three general practitioner contacts, four to seven midwife consultations at regional hospital of birth.	Seven contacts with general practitioner and/or midwife at a public health center.	Nine midwife consultations at regional maternity clinics.
Care at birth	Default of midwife-assisted delivery at regional hospital. Maternity ward care or discharge with follow-up visit.	Default of midwife-assisted delivery at hospital or maternity clinic. Maternity ward care or discharge with follow-up visit.	Default of midwife-assisted delivery at regional hospital. Maternity ward care or discharge with follow-up visit.
Well-child care	Nurse home visiting program for all new families in first year of life (up to five universal visits). Seven preventive health checks at the general practitioner for children aged 0–5, including administration of the vaccination program.	Initial home visit by public health nurse, visits at local child health care centers (midwives, nurses, and physicians monitor children up to age five, 14 visits), vaccination program at health centers.	Initial home visit by public health nurse, visits at regional child health care centers (nurses and physicians monitor children up to age five, 12 visits in first two years), vaccination program at health centers.

Source: Daníelsdóttir and Ingudóttir (2020).

obvious follow-up question to such studies on harmful events early in life is whether there are symmetric positive effects for early-life health policies.

From a theoretical point of view, not only shocks but also public policies such as prenatal care or hospital care at birth are inputs (in addition to family inputs) into the production of good infant health. Models of health and skill formation during different periods of childhood suggest that these policies may have long-lasting impacts on both individuals' health and socioeconomic outcomes over the life course (Cunha and Heckman 2007). These models hypothesize that returns to earlier health investments (during potentially critical ages such as infancy) may be larger than for later ones, and returns to later investments may even be critically dependent on early investments. In this view, early life health policies may lay the basis for individuals being able to benefit from later investments across domains.

Empirically, much of the enthusiasm about the lifelong returns to early-life health policies comes from work on high-intensity policies for at-risk populations, such as specialized health care for very low birthweight infants (Bharadwaj, Løken, and Neilson 2013; Daysal et al. 2022), health insurance and preventive care for at-risk families (Miller and Wherry 2019; Hendren and Sprung-Keyser 2020; Newhouse

2021), or targeted home visiting for disadvantaged families in the United States and elsewhere.² One explanation for why evidence on the effectiveness of universal policies is scarce is methodological challenges, because suitable control groups are typically lacking in universal programs implemented at scale. Moreover, even in the data-rich Nordic countries, often no readily available data on universal programs exist. Similarly, as illustrated in the following sections, easily accessible outcome measures such as mortality are often too coarse when studying universal care.

As a result, the core question of whether findings on the benefits of targeted policies carry over to universal ones remains. From the onset, this question is very much about the relevant counterfactual: Are we concerned about access to universal care (versus no care at all) or do we study variations in universal care provided? In any case, one may expect smaller average impacts from universal programs focusing on an average population than for targeted programs. At the same time, proponents of universal policies emphasize that those may still be worthwhile if they achieve their goals of identifying at-risk populations and initiating relevant care in a timely fashion. In doing so, they may prevent costly specialized treatments downstream and encourage parental health investments with potentially large individual and societal returns. As a result, also for universal care, program effects may be strongest for families with risk factors—both in settings where we study the access to any care or the intensive margin of care.

In the Nordic countries, the origins of today's universal early-life health policies predate the introduction of other well-known and frequently discussed welfare policies, including universal health insurance, maternity and parental leave programs, and universal childcare. Continuing the trend of the nineteenth century, all Western countries saw rapid declines in infant mortality in the opening decades of the twentieth century (Loudon 1992). Many factors—among them public health innovations such as the introduction of sewage management at the turn of the century or the introduction of sulfonamides and antibiotics in the 1930s and 1940s—contributed to this development (Cutler and Miller 2005; Alsan and Goldin 2019; Lazuka 2020; Anderson, Charles, and Rees forthcoming). The Nordic countries (as in the United States at the time) also introduced a set of policies that broadened the general access to trained health professionals (for example, through certified midwives), that increased health monitoring (for example, in maternity wards), and that provided health information (for example, by offering first-year home visits).

Focusing on Denmark, Norway, and Sweden, Table 2 summarizes the main features of today's versions of the early-life health policies offered to a general population of families and discussed in more detail in the following sections. Three commonalities with respect to provision and financing of these policies deserve

²For evidence on targeted home visiting programs, see results from randomized control trials on the Nurse Family Partnership (for example, Olds et al. 1993; Olds et al. 1997, 1998; Eckenrode et al. 2010; Olds 2016). Other targeted programs with randomized control trials include programs in Ireland (Doyle 2020), Germany (Jungmann et al. 2015; Sandner et al. 2018), and Jamaica (Gertler et al. 2014), among others.

mentioning. First, all services are part of the public health care system and accessed by all residents on the same terms. All services are free at time of service and covered under the universal health insurance. Their financing relies on (local and national) government revenues. Some private providers operate within the public programs (such as private prenatal care centers in Sweden) and discussions on their role and greater choice options within the publicly provided care are prevalent in all three countries. Second, national health agencies provide guidelines and regulations (for example, for the suggested number, timing, and content of prenatal care contacts or nurse home visits to new families) and decentralized regional and local public providers offer services to families within the framework of these guidelines. These central guidelines ensure a rather uniform implementation of programs. Third, all services are provided by professionally trained and mostly publicly employed or funded staff, including midwives, nurses, and physicians.

While Table 1 describes the default offer, enrollment in all programs is voluntary. In general, coverage of pregnancy and first-year health programs is very high in the Nordic countries (Daníelsdóttir and Ingudóttir 2020). Using administrative data for all Copenhagen births in 2007–2010, Hirani, Sivertsen, and Wüst (2020) illustrate this broad coverage: for both the number of prenatal midwife consultations (an average of 4.8) and the number of pregnancy checks at the mother's physician (an average of 2.7), the average uptake in their sample corresponds well to the suggested number of contacts. Almost all births take place at hospitals (only 1–2 percent of births are home births). For each of the suggested first-year general practitioner health checks, uptake is around 90 percent, as is uptake of the first-year vaccinations. For nurse home visiting (where no national data exists for any of the Nordic countries at this point), Copenhagen records show that 92 percent of families receive the initial visit after hospital discharge. While consecutive first-year nurse visits display lower take-up of 75–85 percent of families, the average number of universal visits in the sample is 3.3. Additionally, families on average receive one needs-based visit (at the discretion of their family nurse). Of course, this average masks heterogeneity across family characteristics such as the number of children, medical risks, or socioeconomic disadvantage, with first-time parents and families with risk factors typically taking up additional need-based care.

Evidence on Universal Early-Life Health Policies from the Nordic Countries

Prenatal Care

Today all developed countries offer near-universal prenatal care in some form, although pregnant women enter it on different terms (depending on factors such as the provisions of public and private health insurance coverage). In the Nordic countries, universal prenatal care offers regular consultations with midwives and physicians, focusing on information and behavioral advice, as well as monitoring of pregnancy health (with routine procedures such as urine testing and ultrasound

examinations). At-risk families typically receive more intensive follow-ups or referrals to more specialized care. Nearly all families in the Nordic countries enter the public prenatal care program in the first trimester of the pregnancy and the suggested number of prenatal contacts resembles the recommendations in the United States (National Academies of Sciences, Engineering, and Medicine 2020).

Despite its broad coverage, there is no causal research focusing on the introduction of universal prenatal care in the Nordic countries. Some studies, as discussed below, factor in prenatal care components in their analyses. However, it is safe to assume that not the access to any prenatal care, but rather timing and quality, are the most interesting margins for policy discussions in developed countries today. Moreover, while studies examining impacts on maternal and infant health are important, one of the main positive contributions of universal prenatal care may be its impact on parental investment behaviors (such as discouragement of maternal smoking or encouragement of uptake of additional care during pregnancy and after birth).

Only a few Nordic studies have examined these important margins, and those analyses face data and identification problems (in the presence of universal access to the public prenatal care programs for all pregnant women). Two studies confront these issues by exploiting changes in reimbursement schemes and by considering changes in the content of the prenatal care offer. Jensen (2014) exploits a change in the remuneration system for general practitioners in the Danish capital, Copenhagen, in 1987, which introduced a payment scheme relying more heavily on fee-for-service contracts (which were already the norm in other parts of the country). In that system, general practitioners faced incentives to schedule more consultations and perform additional diagnostic tests, including services related to general prenatal care. In a difference-in-differences analysis of reform exposure of prospective mothers, the study finds that birth outcomes improved for exposed pregnancies. Even in this universal setting, however, the effects exclusively come from births to younger women, who may have particular behavioral risks and thus benefit most from low-complexity prenatal care. Those children born under the original fixed remuneration system have slightly lower birth weights (1.8 percent) and a higher probability of being born preterm (58 percent). Unfortunately, the study cannot directly identify the reform impact on the specific type of general practitioner prenatal care, which would be informative about mechanisms—such as pregnancy-related diagnostic tests and consultations—because the national registers did not collect this data. It also remains unclear how much of this health impact at birth matters for longer-run child and maternal health. However, given the finding that only young mothers benefitted from increased intensity of prenatal care, the study points to the importance of differentiating care offers within the existing universal system. This finding is in line with studies from other settings, among them the United States, documenting limited impacts of increasing the number of prenatal care contacts for all pregnancies.

In their study of a combined screening and intervention program within universal prenatal care in Sweden, Grönqvist et al. (2016) directly examine the importance of prenatal care for maternal behaviors. The program targeted maternal

alcohol consumption and expanded over time and across prenatal care clinics. They find that the introduction of structured and systematic screening efforts for all pregnancies (and thus improving identification of at-risk pregnancies), in combination with a focus on counseling about positive health behaviors, improved infant health along selected dimensions (drug prescriptions and hospitalizations in the first year of life but not beyond).³ In an interesting and novel part of their paper, they use survey reports on pregnancy smoking and breastfeeding (in the absence of data on alcohol consumption) to show that treated mothers report significantly higher levels of pregnancy smoking cessation (a 25 percent increase of cessation at the relevant mean of 2.4 percent of mothers who quit smoking during pregnancy). They also find suggestive evidence for a positive impact of the new program component on other investment behaviors (breastfeeding at one month), even though these results are not significant at conventional levels. Thus, their study demonstrates that the quality of universal prenatal care matters for parental investment behaviors and that more research focusing on this margin is warranted.

Evidence on Universal Care at Birth

In the Nordic countries, just as in the United States, the vast majority of births are hospital births. In contrast to the variety of choices in the US system and their close relation to insurance coverage, assignment to public hospitals in the Nordic countries is typically governed by “catchment areas,” where a hospital takes cases from a defined geographic area. The default birth is midwife-assisted, while only risk factors identified prior to or at birth lead to either a physician-assisted birth at a specialized ward or a scheduled Caesarean section.⁴ Care for uncomplicated births in the maternity ward focuses on providing health monitoring for both mother and newborn and on providing information and counseling to new parents (on topics such as infant feeding). Early-discharge and outpatient birth policies have become more common in the past decades for healthy women and their infants, who in this case leave the maternity ward at 4–24 hours after birth and either return for health checks or receive those in the family home.

The lessons learned from research on universal care around birth from the Nordic setting are twofold: First, expanding access to care around birth for broad populations—through skilled midwives and access to maternity wards—has been one important driver of decreases in early life mortality and has had lifelong consequences for treated cohorts. This finding is relevant for similar contexts today and discussions about the importance of the birth setting. Second, alteration of universal care around birth in its core dimensions—the duration of maternity ward care and

³Grönqvist et al. (2017) exploit a decision rule based on the screening tool and find negligible impacts of a follow-up treatment for at-risk women. Without data on take-up, they are unable to determine whether these results are due to limited compliance or lack of effectiveness of the follow-up intervention.

⁴Partly related to underlying health differences across populations, uncomplicated births in the Nordic countries on average are less invasive than in the United States. As an example, for a sample of uncomplicated attempted vaginal deliveries, Maibom et al. (2021) report that 19 percent of births are induced and 24 percent of births receive pain relief through an epidural anesthesia.

the quality of care—can have traceable impacts on average populations of children and mothers.

Several Swedish studies document the importance of access to trained midwives and care at maternity wards: for example, Lazuka (2018) studies the role of certified midwives in rural Sweden at the turn of the twentieth century. For the 1881–1930 cohorts, the share of home births assisted by trained (rather than untrained) midwives increased tremendously, leading to both improved care at birth and in the weeks to follow for mothers and infants.⁵ Trained midwives used antiseptic techniques during birth, knew which procedures to perform (or not) during birth, and cared for mother and infant in the weeks after birth, monitoring their health. In the short run, qualified midwife assistance reduced the risk of neonatal (but not infant and childhood) mortality by as much as 49–61 percent. In the long run, better midwife care around birth provided reduced mortality of exposed individuals at ages 40–80, in particular by reducing the risk of cardiovascular- and diabetes-related death. This latter finding may be attributable to improved infant care and nutrition, as well as prevention of early-life morbidity. Earlier research links both factors to poor adult health.

Sweden was also a frontrunner with respect to large-scale access to maternity ward care at birth (Vallgård 1996). Studying the expansion of maternity wards and access to hospital births in the 1930s and 1940s, Lazuka (forthcoming) documents health and socioeconomic returns to moving the setting of birth-giving. Maternity wards improved hygiene and post-birth care for mothers and infants and facilitated access to drugs, technology, and procedures during birth. Studying the maternity ward expansion across birth years and catchment areas, the study finds that ward openings had effects that cascade through life. Individuals exposed to the expansion were more likely to be actually born in a maternity ward and experienced significant reductions in neonatal mortality, driven by reductions in preventable causes of death. In the longer run, individuals with access to a maternity ward grew up to earn higher labor incomes, as well as to be less likely to leave the labor market on disability pension or be unemployed in the age span from 47–64. These mortality and income effects arise via lower adulthood morbidity and improved educational outcomes of treated cohorts.

More recently, the specific type of general care at birth for healthy populations is frequently altered, but is less frequently studied: While medical care for at-risk births has well-documented benefits, the consequences of changes in (typically less complex) universal care for non-at-risk populations are less clear-cut. One important change in the Nordic setting is the centralization of small maternity wards in larger units. Mergers have both been motivated by the aim of further improving access to specialized health professionals and treatments also for uncomplicated births, as well as cost containment. Policy discussions evolving from mergers center

⁵Work on the Swedish expansion of access to trained midwives in the nineteenth century documents large causal impacts on maternal mortality, but not infant mortality (Lorentzon and Pettersson-Lidbom 2021). In that period, the authors argue, midwives did not provide post-birth infant care.

around the balance and usage of specialized versus general care at birth and around crowding at large maternity wards (which function like emergency care settings and thus have to cope with relatively large variations in case load).

Evaluating the impact of hospital mergers in Sweden, Avdic, Lundborg, and Vikström (2020) find mixed results. They show that ward mergers increase maternal obstetric trauma, suggesting a negative role of crowding and resulting limited access to health professionals such as midwives during birth. At the same time, ward mergers appear to have some positive impacts on infant health at birth (measured as a negative impact on the probability of fetal distress and trauma), driven exclusively by infants with access to higher quality wards after the merger (as intended). Exploiting a difference-in-differences design, Grytten et al. (2014) do not find impacts of hospital mergers in Norway on infant mortality in a low-risk population. However, as discussed, non-mortality margins are likely to be very relevant in this setting. Focusing on Denmark, Maibom et al. (2021) zoom in on crowding by exploiting naturally occurring variation in (day-by-day) admissions to maternity wards, rather than mergers. Thus, they more directly isolate crowding effects for spontaneous births than studies on mergers. They show that on crowded days, prospective mothers are admitted in the ward later (giving them shorter durations of midwife care) and are less likely to receive time-intensive treatments during birth (such as pain relief and inductions, which require access to physicians and midwives). While studies on at-risk births find negative health effects of crowding (Marks and Choi 2019), in the Danish setting and for uncomplicated births, temporary crowding does not appear to induce negative health effects (measured at non-mortality margins, among those diagnosed complications at birth and hospital readmissions). Given downward adjustments of medical procedures like pain relief and inductions on crowded days, these findings may suggest over-provision of those procedures on slow days for uncomplicated births. Taken together, the studies on mergers and crowding highlight some of the potential tradeoffs in decisions on the design of universal care at birth: Alterations of care may have opposite-signed implications by, for example, reducing low complexity care such as midwife access for many and improving access to specialized care for some.

Similarly focusing on the importance of universal care for uncomplicated births, Sievertsen and Wüst (2017) consider how the duration of post-birth maternity care for low-risk mothers and infants affects the health of children and mothers, parental investments, and children's educational outcomes. Motivated by cost containment goals, some Danish regions introduced outpatient birth regulations for uncomplicated births with mothers who had previously had at least one child in the 1990–2003 period. In this setting, healthy mothers and infants leave the hospital early, forgoing monitoring of infant and mother health and timely counseling by health professionals (for example, on topics such as infant feeding). Sievertsen and Wüst (2017) document a strong first-stage effect of mandated early discharge on the probability of experiencing an outpatient birth for mothers who have given birth more than once, but not first-time mothers (as intended by the policy). Both difference-in-differences analyses and a triple difference analysis with first-time mothers

as additional controls show that an outpatient birth increases first-month hospital readmissions (as well as first month general practitioner contacts) for treated infants (a 75 percent increase in infant hospital readmissions for marginal children). While this result may reflect substitution of post-delivery hospital care with readmissions, the study also documents longer-run negative impacts of early hospital discharges on children's educational outcomes: Children who experience an early discharge have a lower ninth-grade grade point average (0.08–0.1 of a standard deviation). Importantly, heterogeneity analyses show that effects for short-run health care take-up are driven by affluent families (who appear to be able to compensate for a lack of universal care), while the longer-run negative impacts are strongest among disadvantaged families.⁶ Pointing to the role of health information and counseling in the maternity care setting for parental investment decisions, their analyses document negative impacts of an outpatient birth for the most disadvantaged families, who are almost 50 percent less likely to report exclusive breastfeeding for four months of the child's life.

These findings about duration of post-birth maternity care, breastfeeding, and test scores from Sievertsen and Wüst (2017) are among the few to link both universal health care, parental investments, and child outcomes across domains. They speak to other recent research outside the Nordic countries with similar conclusions: Fitzsimons and Vera-Hernández (2016) show that low-educated mothers' breastfeeding initiation in the United Kingdom is lower among those who give birth on the weekend rather than during the week. Exploring this observation, Fitzsimons and Vera-Hernández (forthcoming) show that breastfeeding support in the maternity care setting is of lower quality during weekends. Using variation in the day of birth for a sample of uncomplicated births to low-educated mothers, they find that shorter breastfeeding durations for weekend-exposed births have negative consequences for child cognitive outcomes, but not health, in the first five years of life. Finally, the work on outpatient births in Denmark complements Almond and Doyle (2011), who study a similar healthy population of births in the United States. They find no evidence for impacts of an additional night at hospital (rather than just one night) on readmissions or infant mortality. However, they are unable to follow treated children over time and assess other relevant outcomes, such as parental investment behaviors or educational outcomes.⁷ These studies taken together illustrate that factoring in outcomes in the short- and longer-run as well as at non-mortality margins are important for assessing the importance of universally provided care at birth.

Evidence on Universal Well-Infant Care

After hospital discharge, all families in Denmark, Norway, and Sweden can access the well-child preventive care program, consisting of either universal home

⁶Importantly, outpatient birth policies did not affect births with clearly identified health or social risks.

⁷Evans et al. (2008) focuses on mandatory stay laws, which typically serve disadvantaged populations of mothers and children, finding benefits of longer hospital stays in these populations.

visiting and, in parallel, consultations and vaccinations at the family general practitioner (in Denmark), or initial home visiting followed by care at health centers that provide health screening, counseling, and access to vaccinations (in Norway and Sweden). Several randomized control trial studies have shown that targeted home visiting by public health nurses is very effective in improving short- and long-run family outcomes, both in the United States and other settings. The evidence from the Nordic countries—especially from the initial introduction of programs—complements this research and speaks to the benefits of broader home visiting programs.

In the early 1930s, the previously declining infant mortality rate stalled at around 3 percent in Norway and 5–6 percent in Denmark and Sweden (Bütikofer, Løken, and Salvenes 2019; Wüst 2012; Bhalotra, Karlsson, and Nilsson 2017), with a considerable share of mortality being post-neonatal and due to preventable causes related to improper infant care. To decrease infant mortality and to confront demographic concerns related to stagnating fertility rates, the Danish National Health Service initiated the Danish home visiting program, rolled out from 1937 onwards (following a small Danish trial funded by the US Rockefeller Foundation). The program provided around 10 visits in the first year of the child’s life for all families, during which nurses promoted proper infant nutrition (especially breastfeeding), educated parents in infant care, monitored child health and development, and referred ill infants to additional care. Sweden introduced a similar policy in an experiment in the 1931–1933 period in a set of 59 municipalities, selected to be broadly representative for Sweden. The trial introduced home visits and visits at health stations with physician and nurse staff. Cohort eligibility rules in the trial led to differential exposure of mothers and infants with either prenatal care, first-year well-infant care, or a combination of the two. In Norway, well-infant care centers, run by nongovernmental organizations and co-funded with local and state funding, were rolled-out. From 1914 onwards, centers expanded families’ access to health checks, immunizations, health information, and referrals of ill infants to doctors.⁸

The geographic and time variation in exposure to well-infant care together with the data infrastructure in the Nordic countries (combining transcribed historical records and population administrative register data for the cohorts in question) offers a unique opportunity to study the causal short- and long-run impacts of universal well-infant care.⁹ In the short run, studies from all three settings find

⁸Similar policies emerged in neighboring Finland with the roll-out of maternity and well-infant care stations (Krakow 1947; Malin and Hemminki 1992). From 1938 onwards, poor families (and from 1949 onwards all families) received maternity benefits. Those consisted of either an allowance or a “baby box,” providing new families with a basic supply of infant items—with the box serving as the infant’s first bed. While historical accounts emphasize the importance of these policies (Koivu et al. 2020), there is no available quantitative research on their short- or long-run effects. For popular discussions on the topic, see the reporting by Smirnova (2018) and Hakulinen and Gissler (2017).

⁹While none of the described studies factor in recent methodological advances for difference-in-differences with multiple groups and periods (Goodman-Bacon 2021), they all include robustness tests exploring the impact of sample and period restrictions, as well as model specification with respect to trends, controls, and level of aggregation.

important impacts on infant mortality, especially related to preventable causes. For example, exploiting the staggered implementation of the program across Danish towns in the 1937–1949 period together with aggregate historical records, Wüst (2012) documents that infants exposed to the program were more likely to survive their first year of life. Program effects range from 0.5–0.8 percent at the mean infant survival rate (5–8 lives saved per 1,000 live births or a reduction in the average infant mortality rate in treated towns of 4.5 percent by 9–17 percent). Supplementary analyses show that these mortality effects are strong in nutrition-related causes, such as mortality from diarrhea (reductions of 14–23 percent, accounting for between 17–29 percent of the period’s decrease).

Results from Sweden and Norway support these conclusions. Bhalotra, Karlsson, and Nilsson (2017) document for Sweden that trial-exposed infants were 1.6 percentage points less likely to die in the first year of life (a 24 percent decrease at the baseline risk in their sample of 6.6 percent), with infants born to single mothers benefitting the most. Thus, the program contributed to narrowing early health inequalities. Mortality benefits were exclusively driven by individuals exposed to the program after birth, highlighting health monitoring and information to parents about proper infant care as main channels for the effectiveness of the policy.¹⁰ Bütikofer, Løken, and Salvanes (2019) show that the Norwegian intervention decreased infant mortality by 0.8 percentage points for the 1936–1960 cohorts (an 18 percent decrease at a control mean for infant mortality of 4.5 percent). Also in this setting, mortality declines were driven by mortality from relevant causes (primarily diarrhea caused by improper nutrition), which decreased almost 50 percent evaluated at the relevant control mean.

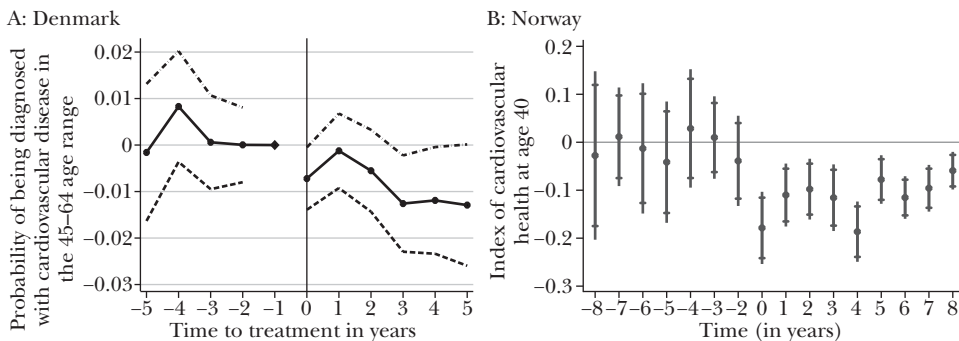
Follow-up studies for the three Nordic programs confirm that initial benefits translated into positive consequences for adulthood health measured both at the mortality and morbidity margin. Across all three countries, studies document positive impacts for adult health issues that have been related to poor early-life nutrition (in line with the short-run results of improvements exactly in nutrition-related dimensions). Hjort, Sølvesten, and Wüst (2017) link information on municipal nurse exposure and individual-level administrative register data in Denmark. They show that individuals of the 1930s and 1940s cohorts, who had access to nurses in their first year of life and were then observed in the administrative data in the 1980–2012 period, are less likely to die in the 45–64 years age range (for example, a 0.2–0.6 percent increase in the survival probability past age 60).¹¹ Moreover, treated

¹⁰ Research from Sweden on the introduction of primary health care centers in the late nineteenth and early twentieth century confirms the positive impacts of universal access to health knowledge and care for infant survival and long-run health, in particular for individuals from disadvantaged backgrounds (Lazuka 2019).

¹¹ In work on the “missing middle” and studying data from military conscription for males born 1940–1949, Wüst et al. (2018) show that the identified long-run health impacts of universal first year care on very long-run health are unlikely to be driven by selective survival into older ages. As this analysis shows no traceable impacts on young adult health and cognitive ability, it supports the hypothesis of a latency period for long-run health effects of early-life interventions. The paper does find a sizeable

Figure 1

Event Study Estimates for the Effect of Exposure to First-Year Home Visiting in Denmark and Infant Center Care in Norway on Later Measures of Cardiovascular Health



Note: Figures are from Hjort, Sølvsten, and Wüst (2017) and Bütikofer, Løken, and Salvanes (2019). The x-axis denotes event time. Point estimates and confidence intervals are from a regression of the relevant outcome measure on event time indicators. For Denmark, the event study further accounts for year and area fixed effects, as well as time-varying area-specific controls. For Norway, the event study includes year and area fixed effects, as well as individual level controls. For Denmark, the outcome measure depicted is the probability of being diagnosed with cardiovascular disease in the 45–64 age range. For Norway, the outcome measure studied is a standardized health index at age 40 summarizing variables relating to body mass index, blood pressure, height, and cardiac and cholesterol risk.

individuals are less likely to have a cardiovascular disease diagnosis in the 45–64 age range (1.3–2.8 percent decrease in the probability of being diagnosed at the control mean of 26.6 percent). Bhalotra, Karlsson, and Nilsson (2017) document large survival returns to exposure to the Swedish trial in the very long run (beyond age 75). Moreover, as for short-run survival benefits, the longer-run benefits in the Swedish study are also larger for socially disadvantaged children. Mortality gains in adulthood are related to causes of death that have been associated with poor infant nutrition and care (among them cardiovascular disease). Similarly, for Norway, Bütikofer, Løken, and Salvanes (2019) find very large improvements (0.3 of a standard deviation) in a standardized adult health index, measuring a set of primarily cardiovascular risk factors at age 40.

Figure 1 presents these results for the impact of early-life exposure to well-infant care on measures related to cardiovascular health for Denmark and Norway in two event graphs. While not directly comparable in magnitude, the patterns for both diagnoses in Denmark and the health index in Norway look remarkably similar, supporting a strong impact of early life care exposure on long-run cardiovascular health. These findings resonate with other work on nutritional shocks and support

positive impact of treatment exposure on the probability of adulthood emigration for men in the sample. As those emigrants on average are positively selected on health and educational status, longer-run results for health and socioeconomic outcomes likely constitute conservative estimates.

programs that has documented the long-run importance of nutrition during infancy and early in life for cardiovascular health in adulthood (Almond and Currie 2011).

While the long-run health results have the described similarities, results for education and labor market outcomes display larger variation: for example, Hjort, Sølvsten, and Wüst (2017) do not find consistent long-run impacts of nurse home visits in Denmark. However, both Swedish and Norwegian studies suggest long-run returns. Bhalotra et al. (forthcoming) combine manually transcribed birth records, primary school records, and administrative data on educational attainment and labor market outcomes to document positive impacts on education and labor market outcomes only for girls in Sweden. Trial-exposed girls grow up to be more likely to complete secondary schooling (17.6 percent), have higher adult earnings (19.5 percent), and have higher full-time employment rates (20.5 percent). The authors attribute the gendered patterns of their results for non-health outcomes to the large concurrent expansion of labor market opportunities for women. In Norway, infant health centers had benefits for the socioeconomic trajectories of treated infants (irrespective of gender), measured as educational attainment and earnings: Infants from municipalities with health centers grew up to complete 0.15 additional years of schooling. Moreover, they experience an increase of their age 31–50 earnings of around 2 percent (Bütikofer, Løken, and Salvanes 2019). Disadvantaged infants benefited more, leading the authors to conclude that the program contributed to a reduction in the intergenerational persistence of educational attainment for the given cohorts.

Given the positive long-run impacts of well-infant care, an important question is its potential interaction with follow-up investments. Empirically assessing the nature of policy interactions across childhood periods is difficult, given the data requirements and the need for two sources of identifying variation. Examining the introduction of nurse home visiting and the parallel expansion of targeted preschools for poor children in Denmark, Rossin-Slater and Wüst (2020) show that both programs had positive long-run health impacts (as found for home visiting in earlier work). Targeted preschools, but not nurse home visiting alone, also improved human capital outcomes of exposed cohorts. In a second step, they show that the positive impact of preschool exposure is smaller for nurse-exposed individuals, indicating considerable substitutability of early life investments that have a strong health component. Their findings underline the importance of more work on the interaction of health and other investment policies over the course of children's lives.

As with the other policies discussed in this essay, well-infant care has changed over time, reflecting societal circumstances and technological changes. In Denmark, for example, from the 1960s on, the focus of home visiting has changed from the prevention of first-year mortality to broader (family) health issues, psycho-social aspects of child development, a coordination of expanding preventive care offers (such as dental care and health checks by school doctors), and childcare. Over time, the number of first-year universal visits has decreased from initially ten to today's five visits alongside other changes. The 1973 mandate of the home visiting

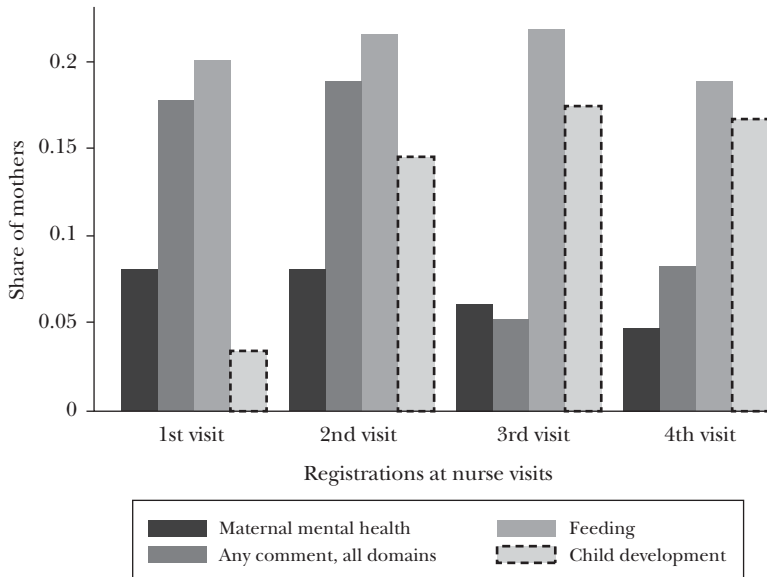
program for all Danish municipalities coincided with births increasingly moving from the home to hospitals (Vallgård 2008). An emphasis on very early home visits after hospital discharge has in recent years evolved in parallel with shorter maternity hospital care, especially for non-first-born children. While some aspects, like monitoring of infant weight gain and counseling on infant nutrition, remain important, today central elements in the contemporary program include aspects such as formal screenings for maternal and paternal postnatal mental health issues. As a consequence, research on the contemporary policy focuses on questions of content, intensity, and duration rather than access to universal care.

Does the timing of first-year well-child care matter? Two papers from Denmark have studied this question by exploiting variation induced by a large-scale nurse strike in 2008. Children born prior to and during this strike missed different elements of universal care: during the strike prenatal midwife visits, hospital care at maternity wards, and post-birth nurse home visits were cancelled at a large scale. Focusing on children born during the two-month strike and comparing their outcomes to children born in the same months of control years, Kronborg, Sievertsen, and Wüst (2016) only find short-lived effects of a reduced care package on take-up of additional health care at general practitioners, their primary measure of child and maternal health. This finding suggests that rather than underlying health effects, the observed increase in general practitioner care in the first month of the child's life constitutes substitution of unavailable default care. Maybe this finding is not surprising given the short duration of the strike and families "bouncing back" to usual care afterwards. However, zooming in on the specific timing of birth of strike-exposed children (early versus later during the strike) the study exploits that families missed different aspects of care: prenatal care, care at birth, or early home visiting. Interestingly, mothers who have children born during the initial weeks of the strike report shorter breastfeeding durations than control families. This finding suggests that especially early nurse visits provide families with health information and counseling that is hard to replace with general practitioner contacts.

In a follow-up study on the same 2008 nurse strike, Hirani, Sivertsen, and Wüst (2020) focus their analysis exclusively on the importance of nurse care during the initial period of family formation. They study families with children who are all born prior to the strike, and thus only vary in the timing of forgone nurse visits in the family home—but not in their access to prenatal care or care in the maternity ward. Strike-exposed families in their sample on average forgo one nurse visit, but they do so at different ages of their child. As another innovation of their study, they examine not only child and mother physical health (measured as uptake of general practitioner care), but also consider maternal mental health. This factor may be instrumental, given documented negative associations of poor maternal mental health on parent-child interactions and child development. Their analyses show that nurse visits in the initial months after birth—as opposed to later ones—affect child and maternal health in the short- and medium-run: that is, early strike-exposed children (and mothers) have more general practitioner contacts during the first

Figure 2

Share of Families with Nurse Registrations across Different Topics at the Four Universal Home Visits



Source: Hirani, Sivertsen, and Wüst (2020)

Note: The figure shows the share of children with a nurse registration in different domains at one of the four universal home visits granted in Copenhagen Municipality. Each domain aggregates a set of binary measures that nurses can make registrations about. “Any comment” is equal to one if the nurse registers any concern across all domains. The sample consists of resident children in Copenhagen born in the period between September 17, 2008, and April 15, 2009.

four years of the child’s life. Moreover, they find evidence for mothers exposed to the strike early in their child’s life facing higher mental health-related risks, even though these results are more suggestive.

These findings raise important questions about mechanisms. Figure 2 from this paper uses nurse records from Copenhagen in non-strike years to illustrate the specific topics covered by nurses during the different universal visits. The figure measures these topics by classifying nurse registrations of specific problems identified at visits. As recommended in the national guidelines, during early nurse visits, nurses screen for (and register) issues relating to maternal mental health. This focus of early visits (and a lack of it for early strike-exposed mothers) may explain the impact of the strike on maternal mental health issues in the longer run. In other words, forgoing an early nurse visit with a mental health screening can have longer-run detrimental mental health consequences. Additionally, Figure 2 illustrates that during early visits, nurses typically cover topics related to infant feeding, a main determinant of infant well-being in the first months. Lacking nurse support and specific information early may directly affect infants’ health and parental

investments. In line with this reasoning, the study documents that the least experienced parents and parents who lack access to health information (first-time parents and parents not educated in a health field) drive their main results for children's health.

Two other studies emphasize the role of health information for all new parents in universal home visits. From January 1992 onwards, Danish home visiting nurses distributed information about changed recommendations on infants sleeping position (from stomach to back), a strong predictor of infant mortality due to sudden infant death syndrome. Altindag, Greve, and Tekin (forthcoming) examine the introduction of new guidelines and their dissemination by home visiting nurses. Using supplementary results from a survey performed at the time, they show that the guidelines were effectively implemented: The rate of infants reported as sleeping on their stomach decreased by 38 percentage points to very low levels of 13 percent around the introduction and dissemination of the new guidelines. In a regression discontinuity design around the release of the new guidelines, the authors demonstrate large and immediate impacts on infant mortality, exclusively driven by sudden infant death syndrome deaths (a reduction of 11.5 deaths per 10,000 live births from a control group mean of 17.8 deaths per 10,000 due to sudden infant death syndrome). As groups with higher pre-determined risks drive these results, such as infants with poor initial health, they conclude that the promotion of health knowledge in the nurse program alleviated health inequalities. This finding indicates that nurses not only provided information, but also tailored their efforts within the universal policy. It further highlights that nurses through their contact with a broad population of parents may play an important role in shaping parental behaviors (like healthy eating habits, prevention of accidents, timely vaccines), which continue to be important for children's health and health inequality.

Finally, studying nurses' role in providing health information, Hirani and Wüst (2022) exploit within-nurse variation in the timing of home visits in a very specific sample of Danish families, who all receive nurse care. They compare vaccination uptake for infants from families who receive their home visit either in the two weeks up to the recommended age for a specific vaccination (treated) or in the two weeks after (control). While parents receive nurse visits as a default offer, they have to schedule vaccination appointments at their general practitioner. The paper asks whether timely nurse visits can impact this parental investment decision. Indeed, a timely nurse visit increases the probability of timely vaccination uptake at three and five months of the child's life. In the longer run, vaccination coverage rates converge between treated and untreated families. However, there is suggestive evidence for inexperienced parents responding to timely nurse visits with longer-run higher levels of vaccination coverage. These findings point to the role of (timely) information provided to families in this setting and suggests, as inexperienced parents show larger and more persistent responses to a timely visit, that nurses not only function as reminders but also may help increase vaccination coverage in specific subgroups.

Conclusions and Future Research

While a variety of studies document the large benefits of early-life investments in at-risk and disadvantaged children, the Nordic countries, with their encompassing set of universal policies, can serve as a natural laboratory to assess the importance of policies that serve larger parts of the population. The introduction of universal early-life health policies in the Nordic countries in the first half of the twentieth century had an important role for improving the health and life chances of treated children. Studies on maternity and well-infant care show important long-run benefits with respect to adult mortality, morbidity, and educational and labor market outcomes and point to their role in explaining the high intergenerational mobility observed for the cohorts in question. In general, the results from long-run follow-up studies underline the importance of factoring in returns over the life cycle when assessing the cost effectiveness of early-life health policies.

Interestingly, the evidence presented here from the Nordic setting echoes results from several studies on US policies of the time, considering both short-run mortality and longer-run health and non-health outcomes. Studying the expansion of access to licensed midwives in the 1900–1940 period, Anderson et al. (2020) show that quality regulations contributed to the decrease in maternal mortality rates, and had moderate impacts on overall infant mortality and strong impacts (a 9.6 percent reduction) on the mortality of nonwhite infants, who were more likely to previously have lacked access to appropriate care. Highlighting the importance of post-natal care, the authors also find evidence for impacts of licensed midwives on the diarrhea-related mortality of children under the age of two. These findings, they argue, may be confounded by other interventions. Moehling and Thomasson (2014) study those potentially confounding policies in their assessment of the 1921 Sheppard-Towner Maternity and Infancy Act, which supported state-level maternal and other well-infant care initiatives. These programs accounted for 9–21 percent of the 1920s decrease in infant mortality, especially for nonwhite infants (0.7–1.9 deaths per 1000 infants). Their findings suggest that spending on personal interactions, such as home visiting by trained nurses, had the largest impacts on infant mortality at estimated costs of about \$20,400 (in 2010 dollars) per infant saved. Similarly, studies focusing on the roll-out of county-level health centers with universal access in US rural areas in the 1908–1930 period document short- and long-run impacts (Hoehn-Velasco 2018, 2021). Centers provided well-infant care by doctors and nurses. Access to this treatment prevented two deaths per 1,000 births, accounting for about 8–10 percent of the mortality decline in the period, most likely driven by improvements in care provided after birth. Impacts were larger in previously underserved areas, and some of the benefits of early-life access to primary care extended into adulthood—as also shown in the Nordic context. Hoehn-Velasco (2021) show that boys exposed to health centers in their first five years of life experienced long-run annual earnings gains of 2–5 percent, and those effects are strongest for early exposed individuals. Treated boys experienced better health rather than increased education, suggesting that improved health in early childhood drives the lifetime benefits of health center exposure.

Causal research on the importance of contemporary early-life health policies in the Nordic countries is still limited and focuses on changes in the organization, timing, and specific content of universal policies rather than access. While this focus makes evidence on contemporary policies appear more scattered, on the whole, existing studies document positive health impacts for children and mothers of providing universal care around birth and of early universal home visits, also in the low-mortality setting. Given the composite nature of the programs studied, pinpointing effective program elements remains a central but empirically challenging endeavor.

Across studies, impacts of universal programs oftentimes are larger for disadvantaged families. This finding is important as in the Nordic countries, concerns about childhood and adult health inequalities have fueled policy discussions about the optimal mix of existing universal policies versus more targeted ones in the future. The results of the studies reviewed here suggest that this topic should be an alley for future research.

The research on prenatal care, maternity care, and well-infant care does not provide easily comparable cost-benefit calculations, and in general those are difficult: Not only would a researcher need to factor in improvements in outcomes across health and socioeconomic domains, over the life-cycle and for both children and their families, but that researcher would also need to answer the question about appropriate counterfactuals. As illustrated in the discussion of the existing research, the margin of “doing nothing” is often not the relevant comparison for policy makers in the Nordic countries. What can be concluded is that the historical introduction of universal policies was highly cost-effective in preventing infant deaths. In Denmark, for example, the estimated costs per averted 1941 infant death through nurse home visiting were around \$10,300 (in 2003 dollars) (Wüst 2012). The per-child costs for providing the policies in the three Nordic countries were very low and fully accrued to the first year of a child’s life (estimated at \$41 in Denmark, \$22 in Norway, and \$39 in Sweden, all in 2014 dollars). Factoring in selected life-time benefits of these programs in terms of gains in (healthy) life-years and benefits related to educational and labor market outcomes for the Swedish and Norwegian policies shows that these policies were effective at very modest costs: for example, stylized calculations for Norway and Sweden suggest internal rates of return of 0.05–0.08 and 0.18 (Bhalotra, Karlsson, and Nilsson 2017; Bütikofer, Løken, and Salvanes 2019).

Today, the contemporary versions of these policies are also provided at modest per child costs. Estimates suggest that Norwegian health care centers provide their default preventive care services at a cost of around \$288 per child. In Denmark, the estimated cost of providing the recommended five universal home visits is \$535 per child (based on figures in Bütikofer, Løken, and Salvanes 2019; Hirani, Sivertsen, and Wüst 2020). However, no encompassing cost benefit analyses factoring in returns across domains and comparing across settings and policies exist. Moreover, conclusions from research on the historical programs emphasize the need to factor in long-run impacts.

Finally, and also important for future research, most existing Nordic studies on early-life universal care still narrowly focus on child and mother health outcomes observed in administrative health care data. This focus neglects important dimensions, such as child cognitive and socio-emotional development or parental stress and postpartum mental health risks. The latter are likely to have longer-run implications for the health and well-being of children and families. Moreover, an important pathway for impacts of universal early-life health policies on children is its impact on parental investments, a point also highlighted in studies on policies for at-risk populations. Too little is known about this pathway. Looking ahead, research should therefore continue to factor in outcomes beyond narrow measures of health care usage, for other members of the family and impacts in the longer-run. In the Nordic countries, much of this work will require researchers to look beyond readily available administrative data, which (despite their high quality) have important blind spots with respect to various aspects of children's health and skill development and investment decisions within the family.

■ *I thank Hans Henrik Sievertsen, Hanna Mühbrad, Timothy Taylor, and the editors Heidi Williams, Nina Pavcnik, and Erik Hurst for very helpful comments and suggestions. I gratefully acknowledge financial support from the Independent Research Fund Denmark, grant 8106-00003B.*

References

- Almond, Douglas, and Janet Currie.** 2011. "Human Capital Development Before Age Five." In *Handbook of Labor Economics*, Vol. 4, edited by O. Ashenfelter and D. Card, 1315–1486. Amsterdam: Elsevier.
- Almond, Douglas, and Joseph J. Doyle.** 2011. "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays." *American Economic Journal: Economic Policy* 3 (3): 1–34.
- Alsan, Marcella, and Claudia Goldin.** 2019. "Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880–1920." *Journal of Political Economy* 127 (2): 586–638.
- Altindag, Onur, Jane Greve, and Erdal Tekin.** Forthcoming. "Public Health Policy at Scale: Impact of a Government-Sponsored Information Campaign on Infant Mortality in Denmark." *Review of Economics and Statistics*.
- Anderson, D. Mark, Ryan Brown, Kerwin Kofi Charles, and Daniel I. Rees.** 2020. "Occupational Licensing and Maternal Health: Evidence from Early Midwifery Laws." *Journal of Political Economy* 128 (11): 4337–83.
- Anderson, D. Mark, Kerwin Kofi Charles, and Daniel I. Rees.** Forthcoming. "Re-Examining the Contribution of Public Health Efforts to the Decline in Urban Mortality." *American Economic Journal: Applied Economics*.
- Avdic, Daniel, Petter Lundborg, and Johan Vikström.** 2020. "How do Closures of Maternity Wards Affect Birth Outcomes?" Unpublished.
- Bhalotra, Sonia, Martin Karlsson, and Therese Nilsson.** 2017. "Infant Health and Longevity: Evidence from a Historical Intervention in Sweden." *Journal of the European Economic Association* 15 (5): 1101–57.

- Bhalotra, Sonia R., Martin Karlsson, Therese Nilsson, and Nina Schwarz.** Forthcoming. “Infant Health, Cognitive Performance and Earnings: Evidence from Inception of the Welfare State in Sweden.” *Review of Economics and Statistics*.
- Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson.** 2013. “Early Life Health Interventions and Academic Achievement.” *American Economic Review* 103 (5): 1862–91.
- Black, Sandra E., Aline Bütikofer, Paul J. Devereux, and Kjell G. Salvanes.** 2019. “This Is Only a Test? Long-Run and Intergenerational Impacts of Prenatal Exposure to Radioactive Fallout.” *Review of Economics and Statistics* 101 (3): 531–46.
- Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes.** 2007. “From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes.” *Quarterly Journal of Economics* 122 (1): 409–39.
- Bütikofer, Aline, Katrine V. Løken, and Kjell G. Salvanes.** 2019. “Infant Health Care and Long-Term Outcomes.” *Review of Economics and Statistics* 101 (2): 341–54.
- Buus, Henriette.** 2001. *Sundhedsplejerskeinstitutionens Dannelse [The Introduction of the Danish Home Visiting Program]*. Copenhagen, Denmark: Museum Tusulanum Press.
- Carneiro, Pedro, Katrine V. Løken, and Kjell G. Salvanes** 2015. “A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children.” *Journal of Political Economy* 123 (2): 365–412.
- Chen, Alice, Emily Oster, and Heidi Williams.** 2016. “Why Is Infant Mortality Higher in the United States than in Europe?” *American Economic Journal: Economic Policy* 8 (2): 89–124.
- Cunha, Flavio, and James Heckman.** 2007. “The Technology of Skill Formation.” *American Economic Review* 97 (2): 31–47.
- Currie, Janet** 2001. “Early Childhood Education Programs.” *Journal of Economic Perspectives* 15 (2): 213–38.
- Currie, Janet, and Maya Rossin-Slater.** 2015. “Early-Life Origins of Life-Cycle Well-Being: Research and Policy Implications.” *Journal of Policy Analysis and Management* 34 (1): 208–42.
- Cutler, David, and Grant Miller.** 2005. “The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States.” *Demography* 42 (1): 1–22.
- Daniélsdóttir, Sigrún, and Jenný Ingudóttir.** 2020. *The First 1000 Days in the Nordic Countries: A Situation Analysis*. Nordic Council of Ministers. <https://pub.norden.org/nord2020-051/#>
- Datta Gupta, Nabanita, and Marianne Simonsen.** 2010. “Non-Cognitive Child Outcomes and Universal High Quality Child Care.” *Journal of Public Economics* 94 (1-2): 30–43.
- Datta Gupta, Nabanita, and Marianne Simonsen.** 2016. “Academic Performance and Type of Early Childhood Care.” *Economics of Education Review* 53 (C): 217–29.
- Daysal, N. Meltem, Marianne Simonsen, Mircea Trandafir, and Sanni Breining.** 2022. “Spillover Effects of Early-Life Medical Interventions.” *Review of Economics and Statistics* 104 (1): 1–16.
- Dietrichson, Jens, Ida Lykke Kristiansen, and Bjørn A. Viinholt.** 2020. “Universal Preschool Programs and Long-Term Child Outcomes: A Systematic Review.” *Journal of Economic Surveys* 34 (5): 1007–1043.
- Doyle, Orla** 2020. “The First 2,000 Days and Child Skills.” *Journal of Political Economy* 128 (6): 2067–122.
- Eckenrode, John, Mary Campa, Dennis W. Luckey, Charles R. Henderson Jr., Robert Cole, Harriet Kitzman, Elizabeth Anson, Kimberly Sidora-Arcoleo, Jane Powers, and David Olds.** 2010. “Long-Term Effects of Prenatal and Infancy Nurse Home Visitation on the Life Course of Youths: 19-Year Follow-up of a Randomized Controlled Trial.” *Archives of Pediatrics and Adolescent Medicine* 164 (1): 9–15.
- Evans, David K., Pamela Jakiela, and Heather A. Knauer.** 2021. “The Impact of Early Childhood Interventions on Mothers.” *Science* 372 (6544): 794–96.
- Evans, William N., Craig Garthwaite, Heng Wei, and NBER.** 2008. “The Impact of Early Discharge Laws on the Health of Newborns.” *Journal of Health Economics* 27 (4): 843–70.
- Fitzsimons, Emla, and Marcos Vera-Hernández.** 2016. “Breast Feeding and the Weekend Effect: An Observational Study.” *BMJ Open* 6 (7).
- Fitzsimons, Emla, and Marcos Vera-Hernández.** Forthcoming. “Breastfeeding and Child Development.” *American Economic Journal: Applied Economics*.
- Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeerch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor.** 2014. “Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica.” *Science* 344 (6187): 998–1001.
- Gomby, Deanna S., Patti L. Culross, and Richard E. Behrman.** 1999. “Home Visiting: Recent Program Evaluations—Analysis and Recommendations.” *Future of Children* 9 (1): 4–26.
- Grönqvist, Erik, Anna Norén, Anna Sjögren, and Helena Svaleryd.** 2016. “Sober Mom, Healthy Baby? Effects of Brief Alcohol Interventions in Swedish Maternity Care.” Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala, Working Paper 2016:16.

- Grönqvist, Erik, Anna Norén, Anna Sjögren, and Helena Svaleryd.** 2017. "Auditing Mothers: The Effect of Targeted Alcohol Prevention on Infant Health and Maternal Behavior." Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala, Working Paper 2017:19.
- Grönqvist, Hans, J. Peter Nilsson, and Per-Olof Robling.** 2020. "Understanding How Low Levels of Early Lead Exposure Affect Children's Life Trajectories." *Journal of Political Economy* 128 (9): 3376–433.
- Goodman-Bacon, Andrew.** 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.
- Grytten, Jostein, Lars Monkerud, Irene Skau, and Rune Sørensen.** 2014. "Regionalization and Local Hospital Closure in Norwegian Maternity Care—The Effect on Neonatal and Infant Mortality." *Health Services Research* 49 (4): 1184–204.
- Hakulinen, Tuovi, and Mika Gissler.** 2017. "Finland's Low Infant Mortality Has Multiple Contributing Factors." *thl Blogi*, January 27. <https://blogi.thl.fi/finlands-low-infant-mortality-has-multiple-contributing-factors/>.
- Havnes, Tarjei, and Magne Mogstad.** 2011. "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes." *American Economic Journal: Economic Policy* 3 (2): 97–129.
- Havnes, Tarjei and Magne Mogstad.** 2015. "Is Universal Child Care Leveling the Playing Field?" *Journal of Public Economics* 127 (C): 100–14.
- Hendren, Nathaniel, and Ben Sprung-Keyser.** 2020. "A Unified Welfare Analysis of Government Policies." *Quarterly Journal of Economics* 135 (3): 1209–318.
- Hirani, Jonas Lau-Jensen, Hans Henrik Sivertsen, and Miriam Wüst.** 2020. "Missing a Nurse Visit." The Center for Economic Behavior and Inequality (CEBI), University of Copenhagen, Working Paper 09/20.
- Hirani, Jonas Cuzulan, and Miriam Wüst.** 2022. "Nurses and Infant Vaccination Coverage." *Journal of Economic Behavior and Organization* 196: 402–428.
- Hjort, Jonas, Mikkel Sølvsten, and Miriam Wüst.** 2017. "Universal Investment in Infants and Long-Run Health: Evidence from Denmark's 1937 Home Visiting Program." *American Economic Journal: Applied Economics* 9 (4): 78–104.
- Hoehn-Velasco, Lauren.** 2018. "Explaining Declines in US Rural Mortality, 1910–1933: The Role of County Health Departments." *Explorations in Economic History* 70: 42–72.
- Hoehn-Velasco, Lauren.** 2021. "The Long-Term Impact of Preventative Public Health Programs." *Economic Journal* 131 (634): 797–826.
- Jensen, V.M.** 2014. "Happy Doctor Makes Happy Baby? Incentivizing Physicians Improves Quality of Prenatal Care." *Review of Economics and Statistics* 96 (5): 838–48.
- Jungmann, Tanja, Tilman Brand, Verena Dähne, Peggy Herrmann, Hüsamettin Günay, Malte Sandner, and Susan Sierau.** 2015. "Comprehensive Evaluation of the Pro Kind Home Visiting Program: A Summary of Results." *Mental Health and Prevention* 3 (3): 89–97.
- Kamerman, Sheila B., and Alfred J. Kahn.** 1993. "Home Health Visiting in Europe." *Future of Children* 3 (3): 39–52.
- Koivu, Annariina, Yen Phan, Ella Näsi, Jad Abuhamed, Brittany L. Perry, Salla Atkins, Mikko Perkiö, and Meri Koivusalo.** 2020. *The Baby Box. Enhancing the Wellbeing of Babies and Mothers around the World.* Helsinki, Finland: Kela.
- Krakow, Samuel.** 1947. "'Neuvola' or Well-Baby Clinic Protects Health of Finnish Children." *The Child* 12 (2): 22.
- Kronborg, Hanne, Hans H. Sievertsen, and Miriam Wüst.** 2016. "Care around Birth, Infant and Mother Health and Maternal Health Investments—Evidence from a Nurse Strike." *Social Science & Medicine* 150 (C): 201–211.
- Lazuka, Volha.** 2018. "The Long-Term Health Benefits of Receiving Treatment from Qualified Midwives at Birth." *Journal of Development Economics* 133 (C): 415–33.
- Lazuka, Volha.** 2019. "Early-Life Assets in Oldest-Old Age: Evidence from Primary Care Reform in Early Twentieth Century Sweden." *Demography* 56 (2): 679–706.
- Lazuka, Volha.** 2020. "Infant Health and Later-Life Labor Market Outcomes: Evidence from the Introduction of Sulfa Antibiotics in Sweden." *Journal of Human Resources* 55 (2): 660–98.
- Lazuka, Volha.** Forthcoming. "It's a Long Walk: Lasting Effects of the Openings of Maternity Wards on Labour Market Performance." *Review of Economics and Statistics*.
- Lorentzon, Louise, and Per Petersson-Lidbom.** 2021. "Midwives and Maternal Mortality: Evidence from a Midwifery Policy Experiment in 19th-Century Sweden." *Journal of the European Economic Association* 19 (4).

- Loudon, Irvine.** 1992. *Death in Childbirth: An International Study of Maternal Care and Maternal Mortality, 1800-1950*. New York: Clarendon Press.
- Maibom, Jonas, Hans H. Sievertsen, Marianne Simonsen, and Miriam Wüst.** 2021. "Maternity Ward Crowding, Procedure Use, and Child Health." *Journal of Health Economics* 75.
- Malin, M., and E. Hemminki.** 1992. "Midwives as Providers of Prenatal Care in Finland—Past and Present." *Women & Health* 18 (4): 17–34.
- Marks, Mindy, and Moonkyung Kate Choi.** 2019. "Baby Boomlets and Baby Health: Hospital Crowdedness, Hospital Spending, and Infant Health." *American Journal of Health Economics* 5 (3): 376–406.
- Miller, Sarah, and Laura R. Wherry.** 2019. "The Long-Term Effects of Early Life Medicaid Coverage." *Journal of Human Resources* 54 (3): 785–824.
- Moehling, Carolyn M., and Melissa A. Thomasson.** 2014. "Saving Babies: The Impact of Public Education Programs on Infant Mortality." *Demography* 51 (2): 367–86.
- National Academies of Sciences, Engineering, and Medicine.** 2020. *Birth Settings in America: Outcomes, Quality, Access, and Choice*. Washington, DC: The National Academies Press.
- Newhouse, Joseph P.** 2021. "An Ounce of Prevention." *Journal of Economic Perspectives* 35 (2): 101–18.
- Nilsson, J. Peter.** 2017. "Alcohol Availability, Prenatal Conditions, and Long-Term Economic Outcomes." *Journal of Political Economy* 125 (4): 1149–207.
- Olds, David.** 2016. "Building Evidence to Improve Maternal and Child Health." *Lancet* 387 (10014): 105–7.
- Olds, David, John Eckenrode, Charles R. Henderson Jr., Harriet Kitzman, Jane Powers, Robert Cole, Kimberly Sidora, Pamela Morris, Lisa M. Pettitt, and Dennis Luckey.** 1997. "Long-Term Effects of Home Visitation on Maternal Life Course and Child Abuse and Neglect: Fifteen-Year Follow-up of a Randomized Trial." *Journal of the American Medical Association* 278 (8): 637–43.
- Olds, David, John Eckenrode, Charles R. Henderson Jr., Harriet Kitzman, Jane Powers, Robert Cole, Kimberly Sidora, Pamela Morris, Lisa M. Pettitt, and Dennis Luckey.** 1998. "Long-Term Effects of Nurse Home Visitation on Children's Criminal and Antisocial Behavior: 15-Year Follow-up of a Randomized Controlled Trial." *Journal of the American Medical Association* 280 (14): 1238–44.
- Olds, David, Charles R. Henderson Jr., C. Phelps, Harriet Kitzman, and C. Hanks.** 1993. "Effect of Prenatal and Infancy Nurse Home Visitation on Government Spending." *Medical Care* 31 (2): 155–174.
- Persson, Petra, and Maya Rossin-Slater.** 2018. "Family Ruptures, Stress, and the Mental Health of the Next Generation." *American Economic Review* 108 (4-5): 1214–52.
- Rasmussen, Astrid Würtz.** 2010. "Increasing the Length of Parents' Birth-Related Leave: The Effect on Children's Long-Term Educational Outcomes." *Labour Economics* 17 (1): 91–100.
- Rossin-Slater, Maya, and Miriam Wüst.** 2020. "What Is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention." *American Economic Journal: Applied Economics* 12 (3): 255–86.
- Sandner, Malte, Thomas Cornelissen, Tanja Jungmann, and Peggy Herrmann.** 2018. "Evaluating the Effects of a Targeted Home Visiting Program on Maternal and Child Health Outcomes." *Journal of Health Economics* 58: 269–83.
- Sievertsen, Hans Henrik, and Miriam Wüst.** 2017. "Discharge on the Day of Birth, Parental Response and Health and Schooling Outcomes." *Journal of Health Economics* 55: 121–38.
- Smirnova, Olga.** 2018. "Why This Country Parcels Babies in Boxes." *BBC*, December 2. <https://www.bbc.com/worklife/article/20181130-why-this-country-parcels-babies-in-boxes>.
- Vallgård, S.** 1996. "Hospitalization of Deliveries: The Change of Place of Birth in Denmark and Sweden from the Late Nineteenth Century to 1970." *Medical History* 40 (2): 173–96.
- Vallgård, S.** 2008. "Widening the Scope, Targeting Interventions, Creating Risk Groups: Maternal and Child Health in Denmark and Sweden from 1930s and Onwards." *Journal of Epidemiology & Community Health* 62 (5): 382–6.
- Wüst, Miriam.** 2012. "Early Interventions and Infant Health: Evidence from the Danish Home Visiting Program." *Labour Economics* 19 (4) 484–95.
- Wüst, Miriam, Erik Lykke Mortensen, Merete Osler, and Thorkild I.A. Sørensen.** 2018. "Universal Infant Health Interventions and Young Adult Outcomes." *Health Economics* 27 (8): 1319–24.

Inequality in Early Care Experienced by US Children

Sarah Flood, Joel McMurry, Aaron Sojourner, and Matthew Wiswall

The care that children experience in early childhood lays the foundation for their future capacities. This paper documents the characteristics of parental and nonparental care experienced by US children before they enter K–12 schooling. We focus primarily on data for children ages zero to four, with occasional discussion of ages five and six. Our analyses span the last two decades, a period during which some important changes occurred in US children’s early care experiences. Care is produced in many different ways: by parents or by other relatives in their own home, in-home by a nonrelative, by a nonrelative in their own home, in public centers such as a school-based pre-kindergarten program or Head Start, or by a private center run by a for-profit or nonprofit organization. Moreover, within each category, care can be provided on a full-time or part-time basis, or in higher- or lower-quality ways.

We build a unified view of children’s care experiences across all types of care to generate new insights into differences in early care experiences between children from families of different socioeconomic status. Our primary finding is that, accounting for both parental and nonparental care, children in families of lower

■ *Sarah Flood is Director of US Surveys, IPUMS, University of Minnesota, Minneapolis, Minnesota. Joel McMurry is a PhD candidate in the Department of Economics, University of Wisconsin, Madison, Wisconsin. Aaron Sojourner is Associate Professor, Carlson School of Management, University of Minnesota, Minneapolis, Minnesota. Matthew Wiswall is Professor of Economics, University of Wisconsin-Madison, Madison, Wisconsin, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are floo0017@umn.edu, mcmurry2@wisc.edu, asojourn@umn.edu, and matt.wiswall@gmail.com.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.199>.

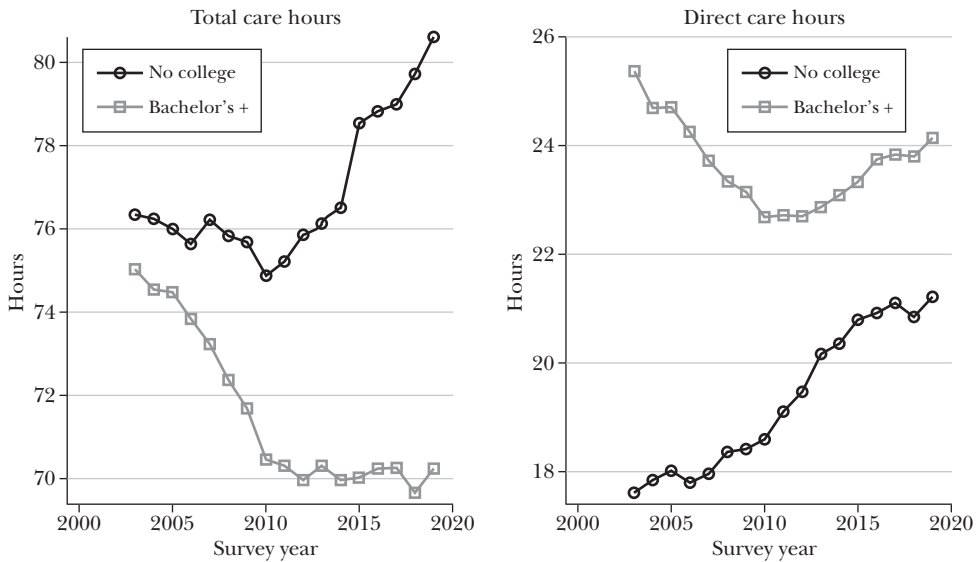
socioeconomic status enter K–12 schooling having experienced substantially lower quality care in all domains during their first five years of life. Gaps are largest for the youngest children and then close as children age.

We also find substantial shifts over the last two decades in care experiences and inequality. As one example, Figure 1 shows evolving patterns of maternal caregiving time. The lines in the figure display average weekly caregiving hours provided by mothers with at least a bachelor's degree and by mothers with no college experience. The left panel shows that in the early 2000s, mothers with more and less formal education averaged similar amounts of maternal care hours weekly, yielding a gap of approximately zero. Since then, mothers with more formal education came to provide fewer hours of care time while mothers with no college education increased the total amount of hours spent with their children, and so the gap widened through 2019. Just prior to the COVID-19 outbreak, higher-educated mothers spent on average about ten fewer hours per week with their children than did mothers with no college experience. Parents of higher socioeconomic status are increasingly specializing in non-childcare activities (primarily labor market work) and making greater use of care from outside their family.¹

The comparisons in Figure 1 illustrate several aspects of our study. We focus on children's *full care experience*, including a range of parental and nonparental care activities, and not on a narrow measure of care such as time parents spent reading to children (Budig and Folbre 2004; Folbre et al. 2005). We focus on a broad range of activities *over two decades*, which allows us to capture substantial changes in care that would have been missed with a shorter time period. We compare across socioeconomic groups, using the level of mother's education as our proxy for child socioeconomic status. We draw upon a variety of datasets: Figure 2, for example, shows that children whose mothers have at least a bachelor's degree are spending fewer hours with their mother but more hours in nonparental care, compared to children of mothers with no college education. The data on care provided directly by mothers is from the American Time Use Survey, while information on nonparental care is from the National Household Education Survey. Overall, our analysis

¹This fact—that parents of higher socioeconomic status spend less total time with their children—may be surprising given prior literature that seems to assert the opposite. In our reading of this prior literature, one issue that causes confusion is the varying meaning of the terms “child care” or “time with children.” For example, an influential paper published in this journal by Guryan, Hurst, and Kearney (2008), focused on a finding from an analysis of 2003–2006 American Time Use Survey data that highly educated mothers spend more time in intensive activities such as reading, playing, and bathing with their children than do other mothers. However, as that paper acknowledges, this subset of intensive-care activities comprises only about one-quarter of the total time that parents spend with their young children, and thus excludes the majority of parental-care time, such as when the parent is eating with the child or is cooking while the child plays nearby. The authors wrote that “high-educated parents and low-educated parents spend nearly identical amounts of total time around their children,” but this point concerning total time seems to have been largely overlooked. Many subsequent papers mistakenly cite the substantial difference in *intensive* parenting time as evidence that parents with more formal education spend more *total* time with their children than do parents with less education. Further, we document that since the 2003–2006 period covered in that earlier study, higher-educated parents spend even *less* total time with their children than do their less-educated peers.

Figure 1

Trends in Mothers' Time Allocation by Maternal Education

Source: American Time Use Survey 2003–2019.

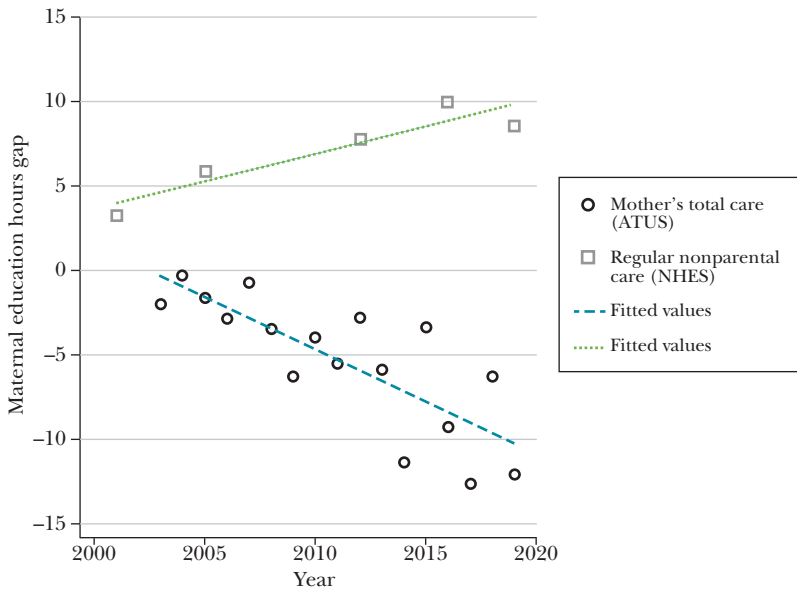
Notes: Sample includes all households with at least one child under the age of four. Five-year moving averages plotted. “Total Care Hours” includes all time the mother is around a child. “Direct Care Hours” is the subset of Total Care Hours in intense childcare activities.

is child-centric where possible, focusing on characterizing the care experienced by young children, rather than on parental time use.

We begin by discussing how we measure concepts such as “socioeconomic status” and “quality of care,” and the data sources upon which we will draw. We then turn to the category of “nonparental care” and see how it varies across children: for example, we document that children from families of lower socioeconomic status receive the majority of their nonparental care from relatives, while most nonparental-care hours for children of households of higher socioeconomic status come from centers or other nonrelative home-based providers. We also evaluate the quality of care provided in these different settings to children from households of differing socioeconomic status, including government programs like Head Start, by looking at detailed measures of activities children engage in with caregivers as well as outside observer ratings of care experiences. A crucial finding is that even within provider type, children from households of lower socioeconomic status receive lower quality care than do those of higher socioeconomic status. Using complementary datasets with measures of parental care quality, we show that children in higher socioeconomic status families also experience higher-quality parental care, reinforcing the advantages of higher-quality nonparental care. Finally, we discuss how to create an overall index of quality of care received by US children, including

Figure 2

Trends in the Gap in Nonparental Hours and in Maternal Care Hours by Socioeconomic Status



Source: American Time Use Survey 2003–2019. National Household Education Survey 2001–2019: nonparental care.

Note: Nonparental is weekly hours in “regular” nonparental arrangements, provided by relatives, centers, or any other nonrelative care providers. “Maternal Education Hours Gap” is the difference between average hours for families in which the mother has at least a bachelor’s degree and average hours for families in which the mother has no college education.

both parental and nonparental care, which reinforces the message that children who grow up in households of lower socioeconomic status receive lower quality of care during early childhood.

Measuring Care

We develop statistics that represent, to the extent possible, the quantity and quality of many varieties of care received by young US children. To measure inequality of care received by children, we need a reliable and standard way to capture socioeconomic status across the various datasets we use. We need ways to capture the quality of care across the range of different settings. And we need to work with a range of data sets. In this section, we explain our process for dealing with these issues, and then present results in the sections that follow.

Measuring Socioeconomic Status

We use the education level of the mother as our proxy for socioeconomic status. Maternal education is well-measured in every dataset we use. Maternal education is also relatively stable. In contrast, children's care experiences, maternal labor supply, and, hence, family income are jointly determined and can shift more frequently. However, our main results are robust to using family income as a measure of socioeconomic status as well.

Maternal education is correlated with other variables commonly used to measure socioeconomic status: educational attainment of others in the family, social network, income, family wealth, and neighborhood, among others. In that sense, it stands as a proxy for this bundle of variables.

Our focus on maternal education also has limitations. Our analyses exclude children living in households without a resident mother. In the National Survey of Early Care and Education data, which we use later in the paper, 7 percent of children under age 5 do not live with a resident mother.

Measuring Care Quality

In measuring the quality of care experienced by children, a significant challenge is that care quality measures tend to differ between parental and nonparental care settings and between different types of nonparental care. All children have the same number of hours per week (T), but each spends this time in a different mix of care settings.

We seek to summarize the quality of a child's overall care experience (Q) and aim to do so in a way that can integrate available information about weekly hours spent in parental care (τ_p) and nonparental care (τ_n) and the qualities of parental care (Q_p) and of nonparental care (Q_n) experienced. To aggregate information on these variables from different sources for the same focal population, we use a weighted average for simplicity:

$$Q = \frac{\tau_p}{T} Q_p + \frac{\tau_n}{T} Q_n.$$

Much of the paper walks through evidence about socioeconomic inequality in each of these variables and concludes with evidence about the magnitude of average care quality differences.

We use age-standardized measures of nonparental and parental quality to measure Q_n and Q_p . In this framework, an aggregate score $Q=0$ implies that a child's overall care experiences are equal to those of the average child, and a score $Q=1$ implies that a child's overall care experiences are one standard deviation higher than the average child's. In this framework, an aggregate quality score of $Q=0$ could be obtained through various combinations of parental and nonparental quality.²

²A limitation of this approach is the implicit equivalence in age-standardized units of our measures of parental and nonparental care. Ideally, we would anchor each separately in some outcome, allowing us to estimate the productivity of each. Chaparro, Sojourner, and Wiswall (2020) estimate this kind of

We consider measures of care quality based on a detailed accounting of activities that parents engage in with children and measures of parental and nonparental quality collected by independent observers. In particular, we focus on “process” measures of nonparental care (measuring the quality of interactions between a caregiver and child), as opposed to “structural” measures (child-to-staff ratios in daycare centers, for example). A well-developed literature has shown that “process” measures are strongly correlated with developmental outcomes, while “structural” measures are less so (Blau 1997, 2000; Early et al. 2006; Mashburn et al. 2008).

Two recent studies also examine quality differences in nonparental care: Pilarz, Lin, and Magnuson (2019) describe how much higher the probability of using any center-based care, proxying for high-quality nonparental care, is for higher-income families and explore other predictors of center use, such as parental employment, nonstandard work hours, and family structure. Bassok et al. (2016) describe variation in quality within and between types of nonparental care (centers, home-based, Head Start) and include analysis of how quality varies with income and other family characteristics.

The current study focuses on measuring care quantities and interactions within different types of care and does not seek to link types of care explicitly to child development outcomes, such as cognitive test scores at kindergarten entry, or even later life outcomes, such as years of completed schooling or adult wages. Of course, such studies require both data that track children over sustained periods of time as well as plausible sources of causal identification. Although we do not focus on later outcomes, a large body of research does estimate strong relationships between early childhood care—both parental and nonparental—and later outcomes (for reviews and recent work, see García, Heckman, and Ronda 2021; Heidlage et al. 2020; Larson et al. 2020; York, Loeb, and Doss 2019; Mayer et al. 2019; Elango et al. 2016; Heckman and Mosso 2014; Shonkoff and Phillips 2000). Thus, while we rely in this essay only on measures of care quality, there is good reason to believe that better quality care translates to better later life outcomes.

Strengths and Limitations of Mixed Data Sources

In making our comparisons, we draw on data from seven primary sources: the National Survey of Early Care and Education (NSECE) (NORC 2012), the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) (US Department of Education 2001–2006), the American Time Use Survey (ATUS) (Hofferth et al. 2020), the Panel Study of Income Dynamics-Child Development Supplement (PSID-CDS)

relationship using experimental data from the late 1980s. In addition, these types of additive indexes implicitly assume that quality in the two types of care are perfect substitutes, up to the respective time weighting. In general, there may in fact be some sort of complementarity across care types in producing particular developmental outcomes (say, measures of cognitive skills at kindergarten entry). In particular, there may be strong complementarities in the sense that very low quality environments can have scarring effects. Caucutt et al. (2020) provide recent estimates of complementarities across several childhood investments. While many other ways to aggregate are possible, almost all would produce a qualitatively similar conclusion.

(Survey Research Center 2019), the National Household Education Survey (NHES) (US Department of Education 2001–2019), the October education supplement of the Current Population Survey (October CPS) (Flood et al. 2021), and the Survey of Income and Program Participation (SIPP) (US Census Bureau 2018). We will briefly discuss each dataset as it is introduced, with more detailed information available in the Appendix.

Here, we just note that each of these data sources has strengths and weaknesses. For example, many studies use the American Time Use Survey, and we do so here as well. But the ATUS covers only parental caregiving hours for one sample parent per household, and thus it cannot deliver an accurate, holistic view of children’s time, nor even just of a child’s time with both parents. Such concerns will be noted as they arise throughout the discussion.

Quality of Nonparental Care

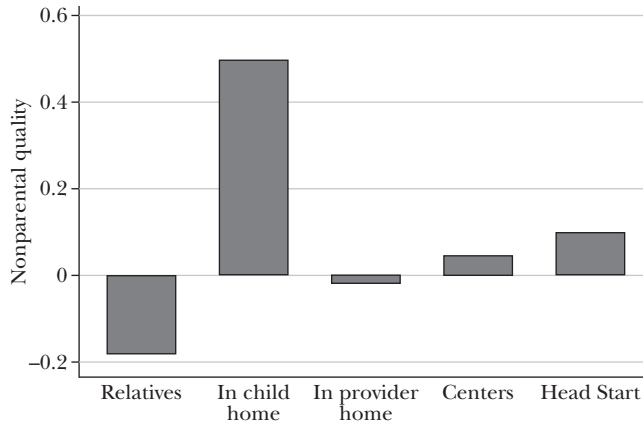
Gaps between Types of Nonparental Care

Children from families of higher socioeconomic status tend to spend larger shares of their nonparental care time in types of care that tend to be higher quality, although these differences diminish as children age into kindergarten. Compared to children whose mothers did not graduate high school, children of mothers with at least a bachelor’s degree spend twice as much of their nonparental care time in public or private centers and about half as much time in the care of relatives during the first three years of their lives. Our measure of quality suggests that centers tend to be about a quarter of a standard deviation higher quality than relative care.

To create a unified view of gaps across levels of socioeconomic status in nonparental care, we harness data on the quality of care by type from the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) together with data on children’s time in different types of care from the National Survey of Early Care and Education (NSECE). From these two datasets, we create a harmonized classification of nonparental care types. We divide nonparental care into five care types, defined by caregiver and the setting in which care occurs: 1) *relative*, which includes grandparents, siblings, or any nonparental relative in any setting; 2) *in child home*, which is a sitter or nanny in the child’s home; 3) *public center*, which includes publicly funded programs such as Head Start, school-based pre-kindergarten, or K–12 school; 4) *private center*, which is a for-profit or nonprofit center without direct public funding, though attending children may receive public vouchers; and 5) *in provider home*, which is care in any other setting, primarily in provider’s home, sometimes referred to as “family child care” or “home-based care.”

In the Early Childhood Longitudinal Study, Birth Cohort data, quality is measured by sending observers to caregiving sites to rate the quality of care against a specified rubric: the Arnett Scale of Caregiver Behavior. This Arnett Scale was designed specifically to be reliable across all types of nonparental care and focuses on caregiver-child interactions. Indeed, the ECLS-B is uniquely useful due to having

Figure 3
Average Quality by Nonparental Care Type (Ages Zero–Four)



Sources: US Department of Education (2001–2006), National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B).

Notes: Average quality measured using Arnett Z-scores. “Centers” are non–Head Start centers.

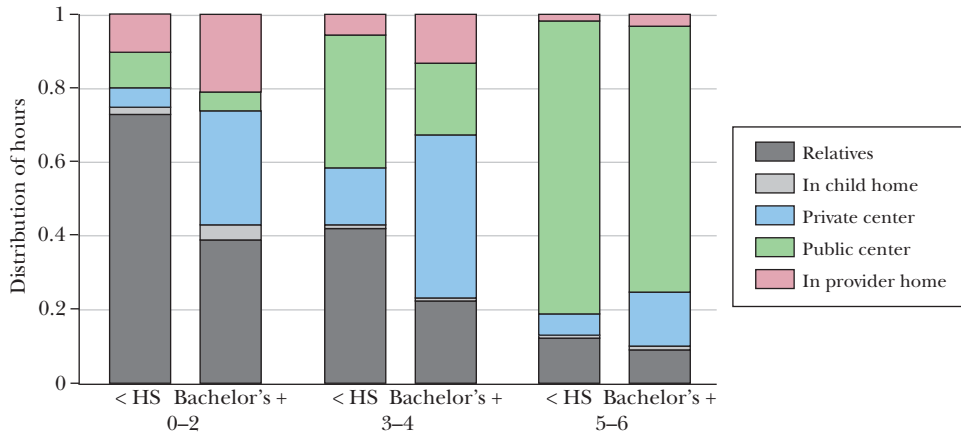
a valid measure of care quality across nonparental care types in a large sample representative of US children. The ECLS-B is a bit dated, representing the cohort born in 2001, but no more-recent data allows this comparison. Other observer-based measures are specific to particular types of nonparental care, such as center care or in-provider home care and so cannot be compared across children with different types of primary providers.

We age-standardize the Arnett scores within the full sample, for each child’s year of age. Figure 3 provides the average standardized Arnett Z-score by care type, for children from birth through age four. The nonparental care type with the lowest average quality is relatives, averaging about 0.2 standard deviations below the mean. The highest-quality type of nonparental care is provided in the child’s home (think “nannies”), at about 0.5 standard deviations above mean. Care in the provider’s home averages just below the sample mean and center-based care tends to score better than average. Here, center care includes any private or public centers, preschools, and pre-K arrangements. The average quality of Head Start care exceeds the average quality of non–Head Start centers, of non–Head Start home-based care, and of relatives. Prior studies report similar rankings (for example, see Bassok et al. 2016; Bernal and Keane 2010).

We can then use the National Survey of Early Care and Education data to examine differences in usage of different types of nonparental care across socioeconomic status. This source provides a careful, nationally representative view of where children spent their time, collected during the first half of 2012. Using child

Figure 4

Distribution of Hours in Nonparental Care Types by Maternal Education



Source: National Survey of Early Care and Education 2012.

Notes: Distribution of hours in “regular” nonparental arrangements. Public center includes Head Start, school-based preschool, and K–12.

and parent time diaries and additional surveys, it aims to capture all sources of nonparental care for each child. For any *regular* (used at least five hours weekly) nonparental providers, information on the provider type is included.

Children from families of lower socioeconomic status receive a far higher share of their nonparental care from relatives than do other children, particularly at younger ages. Figure 4 displays the distribution of nonparental care types by child age group and mother’s education level.³ In the first three years of life, relatives provide about 75 percent of nonparental care for children whose mothers do not have a high school degree but only about 39 percent for children whose mothers have at least a bachelor’s degree. Remember that, as shown in Figure 3, care by relatives tends to be of the lowest quality for child development.

Children from households with a lower level of maternal education also get more of their care from public programs than do children from households of higher socioeconomic status, but the shares of such care and the differences across groups are much smaller than with relative care. In contrast, children from higher socioeconomic status groups get a much larger share of their nonparental care in the first three years of life from private centers and other providers. The share of time with nannies or sitters is very small for both groups, even among those whose mothers have a college degree. The role of public providers (Head Start and school-based pre-kindergarten) expands substantially once children reach preschool age

³To measure public providers, we follow Goerge et al. (2015).

(three to four years old) and the share of care provided by relatives declines. By K–12 age, public programs (kindergarten and first grade) are the dominant form of care for all children.

One might wonder whether these differences across socioeconomic status are explained by differences in employment of mothers across groups. However, when we dig more deeply into the data, this hypothesis does not hold true. Shares of care from different provider types are very similar within maternal education group, whether or not the mother is employed. But as might be expected, for each maternal education group, maternal employment is associated with a larger quantity of nonparental care.⁴

Gaps within Types of Nonparental Care

Quality difference exists within types of care, not just between them. Using the Early Childhood Longitudinal Study, Birth Cohort data, we examine differences in nonparental care quality within care type. Socioeconomic status gaps in the quality of care experienced within care type further reinforce gaps in quality across care types.

Figure 5 plots average Arnett scores of the quality of care by mother's education within broad categories of care types and the child's age group. Children ages zero to two years whose mothers have at least a bachelor's degree experience much higher quality care within care type. The gap in care quality across categories of mother's education is substantial for center-based care (0.4 standard deviations) but even larger still for relative care (0.8 standard deviations). For children ages three through four, we see a similar pattern. The gap for center care is 0.3 standard deviations but over 0.9 standard deviations for relative care. In short, care environments experienced by children in households with high socioeconomic status are above average quality within care type, regardless of type of care. This result echoes Chaudry et al. (2021), who also use the Early Childhood Longitudinal Study, Birth Cohort data but split the sample by family income and use a different measure of quality.

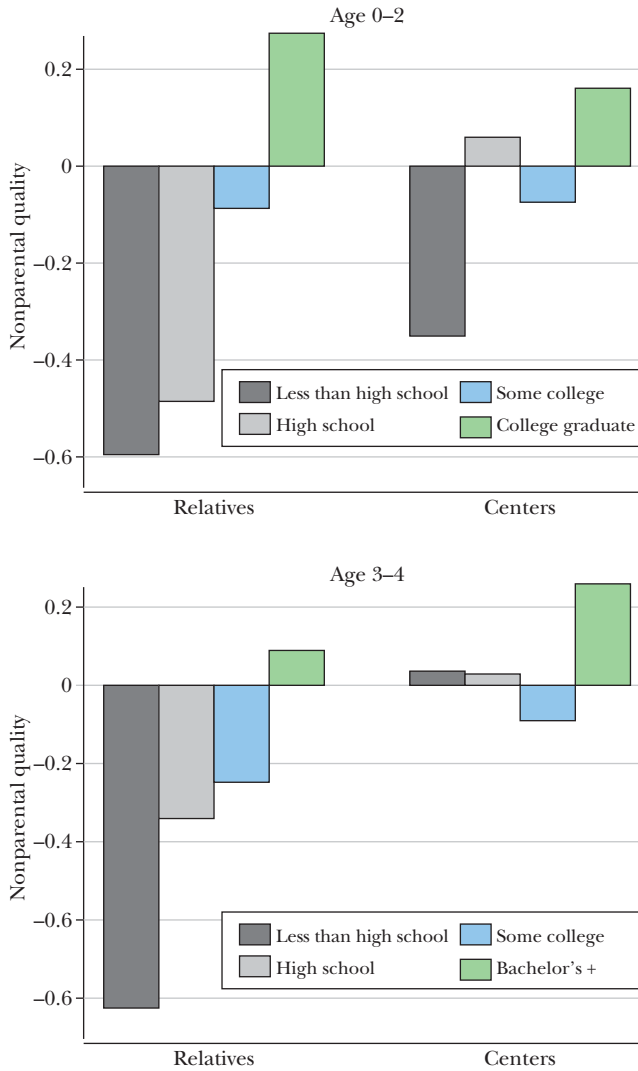
The gaps in quality of nonparental care across and within care types combine to produce gaps in the overall quality of non-parental care experienced by children of different socioeconomic status. The darker bars in Figure 5 present average quality for the primary nonparental care arrangement, regardless of type, within maternal education level but aggregating across ages. They reveal a roughly 0.4 standard deviation difference in nonparental care quality between children of mothers with a bachelor's and those without a high school degree.

Head Start and Quality of Care

In closer examination of this data, we also find that the quality gap in nonparental care quality gets smaller as children age, because children ages three through four from households of lower socioeconomic status experience a higher proportion

⁴The online Appendix offers data and illustrative figures on this point: see A-8 for the zero-to-two age group, and A-20 for ages three to four.

Figure 5
Average Quality of Nonparental Care by Mother Education, Child Age, and Care Type



Sources: US Department of Education (2001–2006), National Center for Education Statistics, and Early Childhood Longitudinal Study, Birth Cohort (ECLS-B).

Notes: Average quality measured using Arnett Z-scores. “Centers” is all centers, including Head Start.

of center-based and Head Start care, which tends to be of higher quality than the other kinds of nonparental care they commonly receive.⁵

⁵Although the price of care is not a direct measure of quality, if one makes the plausible assumption that price is correlated with quality, it can provide another measure. Using National Survey of Early Care and

Indeed, we can also use this framework to explore the quantitative importance of a leading government childcare program, Head Start, in raising the quality of care experienced by children from low socioeconomic status households and, thus, reducing the socioeconomic gap in nonparental care quality. The federal Head Start program provides direct funding primarily for childcare centers to provide free (or highly subsidized) care to eligible children, primarily three and four years old, from low-income households. About 25 percent of the Head Start participants were younger than age three and served by Early Head Start (ACF 2021). In 2019, Head Start served about 1 million children, at an outlay of about \$10 billion in federal government funding. This level of funding allows it to serve 11 percent of eligible children under age three and 36 percent of eligible of three and four year-olds (NHS 2020).

Despite Head Start serving far less than the majority of eligible children, the program plays a large role in lifting the average quality of nonparental care quality for disadvantaged American children.⁶ The lighter bars in Figure 6 present average quality of care excluding households who report using a Head Start care arrangement. The importance of Head Start is evident, especially among the children from lower socioeconomic status households. Excluding Head Start pushes the average nonparental care quality for children in low socioeconomic status households down by more than 0.2 standard deviations, doubling the disadvantage relative to the sample mean.⁷

This rough calculation does not necessarily serve as the policy counterfactual effect of defunding Head Start. In the absence of the program, care patterns would shift in ways that are difficult to foresee. The absence of Head Start would also have broader implications for children's health and well-being. In addition to care, Head Start provides an array of services to enrolled children and their families, including health and home visiting services.

Quality of Parental Care

To this point, our discussion has shown that children from families of higher socioeconomic status tend to experience more time in nonparental care and

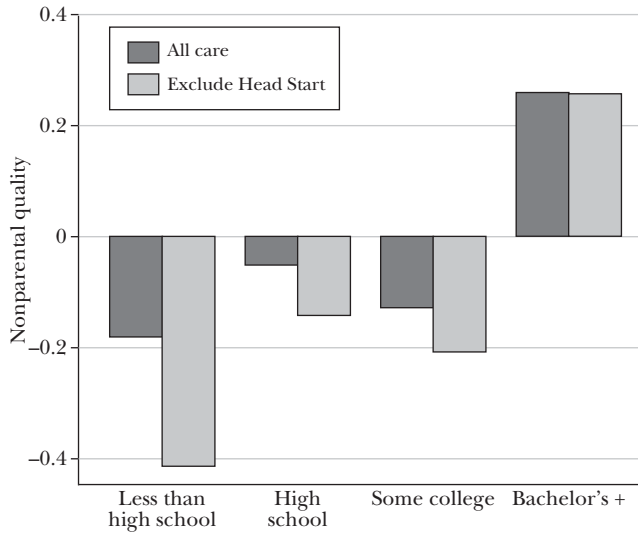
Education data, Appendix Figure A-23 displays median and mean price *paid* per hour for center-based care, conditioning on maternal education and child age. The center-based care that children whose mothers have a bachelor's degree experience is significantly more expensive per hour than that of children whose mother do not have a high school degree: the ratio of average prices paid is roughly three to one. Of course, this difference is certainly attributable to a variety of factors—including location, price discrimination, differences in the valuation of nonquality attributes, and the effectiveness of search—but it is also supportive of and consistent with the evidence that children of mothers with a bachelor's degree experience higher quality center-based care.

⁶Among families eligible for public childcare subsidies, the quality of care received from Head Start tends to exceed that received either using Child Care and Development Fund vouchers or in unsubsidized care (Johnson, Ryan, and Brooks-Gunn 2012).

⁷Appendix Figure A-24 displays the analogous statistics conditioning on family income instead of mother's education as a proxy for socioeconomic status. As discussed earlier, the same conclusion holds.

Figure 6

Average Nonparental Care Quality with and without Head Start, by Mother Education



Source: Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B).

Notes: Average quality measured using Arnett Z-scores. “Exclude Head Start” uses only non-Head Start arrangements.

higher-quality nonparental care. We now show that differences in the quality of parental care tend to reinforce this gap. For every measure of parental care quality studied, children from families of higher socioeconomic status experience substantially higher quality parental care on average.

In this analysis, we draw on three types of data to measure quality of care: parent time diaries from the American Time Use Survey, which provide information about the time parents spend with their children; child time diaries from the Panel Study of Income Dynamics-Child Development Supplement (PSID-CDS), which describe the activities children experience, the number of parents present, and the intensity with which parents participate in the activity; and outside-observer based measures of parental engagement and the home environment, which are available in the PSID-CDS.

With the American Time Use Survey and Panel Study of Income Dynamics-Child Development Supplement time diary data, we use the available variables to construct comprehensive measures of the total time that parents spend with their children (ATUS) and the total time that children are in the presence of their parents (PSID-CDS). Within each of these measures of total parenting time (parental time “given” and “received,” respectively) we classify how much of this time is spent in intense parental activities (for example, reading or playing directly with a child

rather than the parent working in one room and the child playing in another). We can further classify parental time spent in certain activities, including educational activities (like reading or being read to, or receiving music lessons) or screen-based leisure time (television or computers). We focus on the share, rather than the level, of parental care time in these categories because we aim to measure the quality of children's overall care experiences. Focusing just on the level of parent time in intense or direct care conflates the quantity of parental care time with its quality and obscures the fact that children have a time budget. Any time spent in parental care crowds out nonparental care time, which has its own quality level. The major difference between the two data sources is that the ATUS provides parent-level time diaries and the PSID-CDS provides child-level time diaries (completed with the help of parents, of course) and other child-specific measures.⁸

Many prior studies of the quality of parent care focused on parent time diaries from the American Time Use Survey. This data has substantial advantages, including a large sample size and annual surveys covering the past 20 years. However, the ATUS has limitations for our purposes: this survey only records the total time allocation of one respondent parent and does not allow an assignment of time to specific children in the household. As a result, maternal education differences in these data may not necessarily reflect the gradient in care received by children if parents allocate care among their children differently across the education spectrum. However, we find that conclusions drawn from the ATUS are consistent with those from the Panel Study of Income Dynamics-Child Development Supplement, which allows for a child-specific assignment of care time.

Three main conclusions emerge from this exercise. First, from the American Time Use Survey, highly-educated mothers spend a larger fraction of their parental time in the kinds of activities that are considered more likely to promote child development. Using the sample of households with at least one child under age five, pooling the most recent ATUS surveys from 2016–2019, and conditioning on a full set of indicators for the number and ages of household children, we find that while mothers with a bachelor's degree spend about 14 *fewer* hours in total childcare than do mothers without a high school degree, these mothers spend nearly six hours *more* in direct, intensive childcare per week (for details see Appendix Table A-3). This implies that the average fraction of parenting time spent in more intensive activities is substantially higher for mothers with a bachelor's degree, over 70 percent higher. We see similar patterns for father's time for these same households: while total childcare time is about the same, fathers in highly educated households (mother has a bachelor's degree) devote about five more hours per week in direct, intensive childcare than do their less-advantaged peers, with a similarly larger fraction of their parenting time spent in these activities.

Second, turning to the Panel Study of Income Dynamics-Child Development Supplement child time-diary data, we see patterns consistent with the American

⁸For details on the classification of activities, like how night sleep is handled and other matters, see Appendix Section B.3.

Time Use Survey parent diaries. More-advantaged children spend a higher fraction of their total time with parents receiving intense and active parenting (about 5–6 percentage points higher, representing 10–20 percent more relative to children of low socioeconomic status households), a higher fraction of time in education activities (about 2 percentage points higher, representing about two-thirds more), and a smaller fraction of their parental time in screen-based leisure time (about 4 percentage points lower, representing about one-quarter less). Further, when receiving parental care time, children of highly educated mothers are more likely to be cared for by two parents (8 percentage points more likely, representing a quarter more relative to children of low socioeconomic status households) and by the father only (almost 6 percentage points, representing three-quarters more). Their less-advantaged peers spend a larger fraction of parental time with the mother only. However, we note that using which parents are present during parental time received has shortcomings: it omits nonparent caregivers present and the number of other children present.⁹

Third, the Panel Study of Income Dynamics-Child Development Supplement data contains several additional measures of parental care and home environment characteristics, from both parents' self-reports and from direct observation of parent-child interactions and the home environment by trained observers during a home visit. With these, we build additional measures of parental-care quality. Our primary measures of parent and home environment quality come from the Home Observation for Measurement of the Environment (HOME) Inventory. The HOME Inventory is a set of items that aim to measure the quality of a child's home environment, including interactions with the primary caregiver (almost always the mother). We partition the HOME items into three sets: observer-based measures of the primary caregiver and interactions between the caregiver and child, observer-based measures of the home environment, and parent-reported survey items. The Data Appendix describes how we construct Z-score indices from the HOME score items. By each measure, children from higher socioeconomic status households have higher quality care experiences at home. For example, when pooling all items of the HOME inventory, children of mothers with a bachelor's have an average score of over 0.8 standard deviations higher than that of those with no college education.

We conclude our analysis of parental quality patterns by noting that restricting focus to a small subset of intense parental care activities, such as parents reading to children, can suggest that there has been a trend toward convergence in parental care experiences over the last two decades; for example, Cha and Park (2021) and Prickett and Augustine (2021) show a “converging” trend in some intense parenting activities using the American Time Use Survey data. Although children from more-advantaged households do continue to receive more hours of this intense parental

⁹These statistics from PSID-CDS are based on regressions with controls for child age and survey years fixed effect, available in the online Appendix (see Table A-1). In addition, Appendix Table A-9 contains analogous estimates including additional control variables, and Appendix Table A-10 presents estimates using levels as the outcome variable rather than shares.

engagement at least by 2019 (the last year of our data, pre-COVID-19), the gap has fallen, as shown in Figure 1. However, the time that parents from households of higher socioeconomic status spend with their children continues to be of much higher average quality as measured by the fraction of that time in intense or education-specific activities, and other observer-based measures.

Quantities of Care

Our next step is to quantify the differences across socioeconomic status in the amounts and shares of child time spent in parental and nonparental care. We establish that children from households of higher socioeconomic status spend less of their time in parental care than do less-advantaged children. Then, to construct a measure of each child's aggregate care quality that combines information on the quantities and qualities of parental and nonparental care, we turn to the Early Childhood Longitudinal Study, Birth Cohort dataset. While the datasets used above have superior measures of parental-care quality and time in (non)parental care, the ECLS-B is uniquely valuable here because it contains measures of the joint distribution of parental and nonparental care quality, and time in each. Crucially, the ECLS-B contains measures of nonparental care quality that are comparable across nonparental care types.

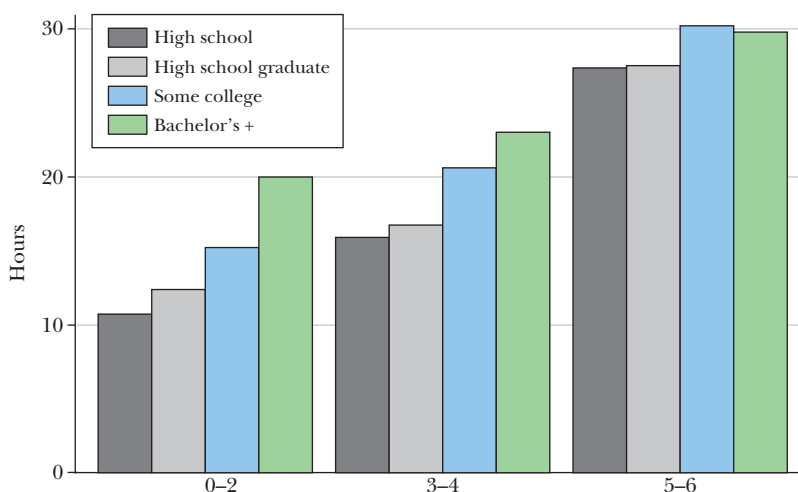
Among all young children, weekly nonparental care hours increase as children age, but they do so especially quickly as children age into eligibility for more public care and education programs. Our discussion of hours of nonparental care draws upon the 2012 child time diaries from the National Survey of Early Care and Education, which provide a comprehensive, nationally representative, relatively recent view of children's time, permitting the best measure of parental versus nonparental care available, overall and by age and socioeconomic status.¹⁰

In their first two years of life, children average about 12 hours of nonparental care weekly. These averages mask large differences in nonparental care hours by socioeconomic status. For the youngest children (ages zero to two), those children whose mothers have less than a high school degree receive just over 10 hours of nonparental care; for children whose mothers have a high school degree, 12 hours; for some college, 15 hours; and for children whose mothers have at least bachelor's degree, nearly 20 hours weekly of nonparental care (Figure 7).

As children reach ages three and four, the amount of time they spend in nonparental care rises for all levels of socioeconomic status. However, the rise is faster for those whose mothers have less education, and thus the differences across groups diminish. Children three to four years old whose mothers have less than a high

¹⁰For figures and additional numerical detail on the patterns of hours of nonparental care discussed here, see online Appendix Section A. One limitation worth noting here is that the survey took place in the winter and spring, and so it does not represent children's experiences in the summer or fall. It is important to note that the statistics for average hours include households that report zero hours.

Figure 7

Average Weekly Nonparental Care Hours by Child Age and Maternal Education

Source: National Survey of Early Care and Education 2012.

Note: Comprehensive nonparental care hours for each child.

school degree average about 15 hours of nonparental care per week (up from just over ten hours for the zero to two year-olds). Children whose mothers have at least a bachelor's average roughly 22 hours weekly (up slightly from nearly 20 hours for the zero to two year-olds).

By the time children reach ages five and six, when almost all children are in K–12 schooling, nonparental care rises to an average of about 30 hours per week. At this point, the gradient by socioeconomic status in the time that children receive nonparental care essentially disappears.

These differences are also apparent if one looks at the share of children receiving full-time nonparental care, which we define as 31 or more hours per week. In the first years of life, fewer than 20 percent of children whose mothers do not have a high school diploma receive full-time nonparental care (see Appendix Figure A-11). At the same ages, the share of children receiving full-time care among those whose mothers have a bachelor's degree is about twice as high.

The other datasets examined tell a similar story about the evolution of nonparental and parental care for recent years. The PSID-CDS provides child time diaries and tells a very similar story. The October Current Population Survey shows a similar gradient for children age three and up using a more-limited subset of nonparental care types: “pre-school,” “nursery school,” or “pre-K.” Similarly, recent data from the Survey of Income and Program Participation show comparable educational gradients in the incidence of using any nonparent care and nonparent, nonrelative care. Details for these alternative measures are available in the online Appendix.

Average Care Quality Experienced by Children

Figure 8 plots the gaps in average care quality experienced between children with four-year-college-educated mothers versus those whose mothers have a high school degree or less. For this illustrative calculation, we omit the “some college” group entirely. We construct this figure using the Early Childhood Longitudinal Study, Birth Cohort data, computing the overall care quality for each sample child, and using the age-standardized frequency parents read to their child per hour of parental care, the age-standardized Arnett scores, and the number of weekly hours in nonparental and parental care to measure parental quality, nonparental quality, and time in each (Q_p , Q_n , τ_n , and τ_p), respectively.¹¹

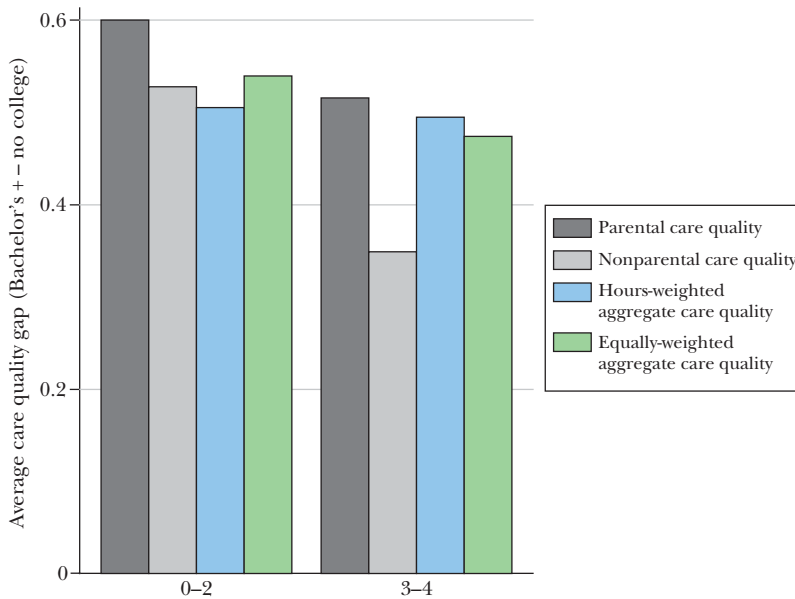
Analysis of these child-level indices of aggregate care quality reveals several patterns. First, those children whose mothers have a higher level of education receive better quality care in both parental and nonparental settings at all ages during early childhood, as shown by the first bars for each of the age groups. Second, at both age groups (zero to two and three to four), the gap in age-standardized parental care quality is largest, and incorporating the quality of nonparental care tends to moderate inequality, particularly for younger children. Third, for the third and fourth bars shown for the two age groups of Figure 8, we also construct overall measures of aggregate quality weighted by hours spent in parental and nonparental care, and an equal-weighted measure in which the quality scored for both parental and nonparental care are given a weight of 0.5. We find a similar gap as using the hours-weighted measure, indicating that hours allocated across care types are not greatly influencing the pattern of overall inequality.

An aggregate care quality index allows a quantitative assessment of the importance of two kinds of potential compensating behaviors: first, parents choosing higher-quality care in one domain to compensate for lower-quality care in the other domain, and, second, parents choosing low numbers of hours of the lower-quality care in favor of more hours of the higher-quality care. We conclude that aggregate measures incorporating both parental and nonparental care experiences also show substantial inequality in care experienced, with a larger degree of inequality for the youngest children.

It is possible that inequality of care obtains in both nonparental and parental domains, but is mitigated when considering aggregate care experienced. We investigated and ruled out two channels through which this may happen: 1) the possibility that children who experience higher-quality parental care tend to experience lower-quality nonparental care, and vice versa; and 2) the possibility of dynamic compensating behavior, whereby higher quality care experiences during preschool

¹¹ To measure parental quality per hour of care, we divide reading frequency by total parental care hours. Reading frequency is recorded as the midpoint of each categorical bin, with 0 recorded for “never,” 1.5 for “1–2 times,” and 4.5 for “3–6 times,” and 7 for “every day.” Frequency per hour is then age-standardized for each child age (in years) to create a Z-score (mean 0 and standard deviation 1 by construction) measure of Q_p .

Figure 8

Parental, Nonparental, and Aggregate Care Quality

Source: Department of Education, National Center for Education Statistics, and Early Childhood Longitudinal Study, Birth Cohort (ECLS-B).

Note: The vertical axis measures average age-standardized (z-score) difference in quality. Gap of 0 represents no difference, gap of 1 represents 1 standard deviation difference. Parental care quality measured using frequency parent reads to child. Nonparental care quality measured using Arnett score for primary nonparental care arrangement.

age might compensate for lower quality experiences as infants and toddlers, and vice versa.

In considering the first possibility, we find that children who tend to experience higher-quality parental care also tend to experience higher-quality nonparental care. We use the Early Childhood Longitudinal Study, Birth Cohort data to provide evidence on the joint distribution of nonparental and parental care quality as measured by the age-standardized Arnett score in the primary nonparental care arrangement against parental care quality with the commonly used measure of reported frequency of reading to the child.¹² For younger (age zero to two) and older (three through four) children, quality of nonparental care is on average higher for children whose parents read to them more, with the gradient

¹²We adapt a commonly used parental activity measure, the number of times parents report reading to their child over a week (for example, Cunha, Heckman, and Schennach 2010; Todd and Wolpin 2007; Blau 1999). For evidence on the developmental effects of reading, see Demir-Lira et al. 2019, Price and Kalil 2019, and the studies cited therein.

more pronounced for younger children. In a series of regressions, we find a positive correlation in measures of parental and nonparental quality, both overall and within sub-samples by mother's education (Appendix Table A-4).

For additional evidence on the correlation of parental and nonparental care quality, we turned to data from the American Time Use Survey. Although the ATUS data do not provide any information on nonparental care, as we have documented from other data, children from less-advantaged families tend to receive lower quality nonparental care. Do their parents compensate for this through higher-quality, more intensive care at home? Rather than seeing evidence that parental care is compensating for lower-average nonparental care, we see evidence of reinforcement. Average parenting quality, as measured using the fraction of time in intensive parenting, is lowest in families in which the mother provides the most hours of care, and highest when families rely on more nonparental care (see Appendix Figure A-1).

Concerning the second possibility of dynamic compensating behavior, we explore the persistence of inequality in care quality. Specifically, we exploit the longitudinal data in the Early Childhood Longitudinal Study, Birth Cohort to regress the quality of care in the later wave of the Early Childhood Longitudinal Study-Birth Cohort (period $t + 1$, when most children are between 3 and 4 years of age) on the quality of care in the earlier wave of the ECLS-B (period t , when most children are between 1.5 and 3 years of age). We see a statistically significant positive correlation between measures of quality of care as children get older (Appendix Table A-5). Similar, and in many cases even higher, levels of persistence are evident for parental care quality and overall quality constructed using the hours-weighted index described above. Although these findings do not indicate perfect persistence (with a caveat that classical measurement error in any of these measures would tend to attenuate findings toward zero), these results do seem to rule out strong dynamic compensating behavior, whereby higher-quality care experiences later compensate for lower-quality experiences earlier.

Conclusion

Prior to age six, at which age almost all American children are enrolled in public or private schooling, children from households of higher socioeconomic status, as measured by the mother's level of education, experience higher-quality nonparental and parental care. In particular, children from families of lower socioeconomic status experience lower-quality care because: 1) they spend more time in the types of nonparental care that tend to be lower quality on average; 2) they tend to experience nonparental care that is lower quality for its type; 3) they tend to experience lower-quality care when with parents; 4) a child's parental and nonparental care qualities tend to be positively correlated such that deficits in one tend to be reinforced in the other; 5) they spend more time with parents and the gap in care quality across socioeconomic status is larger in parental care than in nonparental

care; and 6) quality levels persist across early childhood such that those receiving lower-quality care earlier also tend to receive lower-quality care later. We want to put to rest the common, yet mistaken, claim that more-educated parents spend more total care time with their children than do less-educated parents. However, it is true that they spend more parental care time in narrow sets of activities believed to accelerate child development.

As noted at the beginning of this paper, we focus here on providing evidence for these differences in the quality of care received by children across socioeconomic status. However, we defer to the existing experimental and quasi-experimental evidence to demonstrate causal relationships between measures of quality and child outcomes. Notable in particular is evidence that children who experience poor quality care in and out of the home (“double trouble”) experience worse outcomes (Watamura et al. 2011).

Understanding the differential patterns of care received by children is central to designing policies to address inequality in early childhood opportunity. Across all levels of government, American communities annually invest only about \$1,500 per child in care and education during ages zero through four but nine times that amount in each of the following 13 years of life (Davis and Sojourner 2021). Although there is unlikely to be any single policy solution, improving care for young children in households of lower socioeconomic status likely requires expanded programs for government support of high-quality nonparental care (Hotz and Wiswall 2019; Davis and Sojourner 2021) as well as support for higher-quality parental care (Cunha, Elo, and Culhane 2020; Maloney et al. 2015; York, Loeb, and Doss 2019; Mayer et al. 2019).

■ *We thank David Blau, Jessica Brown, Liz Davis, Daniela Del Boca, Sandy Hoeffferth, Lance Lochner, Taryn Morrissey, Alejandra Ros Pilarz, Anson Zhou, and the JEP editors for helpful comments.*

References

- Bassok, Daphna, Maria Fitzpatrick, Erica Greenberg, and Susanna Loeb.** 2016. “Within- and Between-Sector Quality Differences in Early Childhood Education and Care.” *Child Development* 87 (5): 1627–45.
- Bernal, Raquel, and Michael P. Keane.** 2010. “Quasi-Structural Estimation of a Model of Childcare Choices and Child Cognitive Ability Production.” *Journal of Econometrics* 156 (1): 164–89.
- Blau, David. M.** 1999. “The Effect of Child Care Characteristics on Child Development.” *Journal of Human Resources* 34 (4): 786–822.
- Budig, Michelle J., and Nancy Folbre.** 2004. “Activity, Proximity, or Responsibility? Measuring Parental

- Childcare Time.” In *Family Time: The Social Organization of Care*, edited by Nancy Fobre and Michael Bittman, 51–68. London: Routledge.
- Caucutt, Elizabeth M., Lance Lochner, Joseph Mullins, and Youngmin Park.** 2020. “Child Skill Production: Accounting for Parental and Market-Based Time and Goods Investments.” Unpublished.
- Cha, Yun, and Hyunjoon Park.** 2021. “Converging Educational Differences in Parents’ Time Use in Developmental Child Care.” *Journal of Marriage and Family* 83: 769–85.
- Chaparro, Juan, Aaron Sojourner, and Matthew Wiswall.** 2020. “Early Childhood Care and Cognitive Development.” NBER Working Paper 26813.
- Chaudry, Ajay, Taryn Morrissey, Christina Weiland, and Hirokazu Yoshikawa.** 2021. *Cradle to Kindergarten: A New Plan to Combat Inequality*. New York: Russell Sage Foundation.
- Cunha, Flávio, Irma Elo, and Jennifer Culhane.** 2020. “Maternal Subjective Expectations about the Technology of Skill Formation Predict Investments in Children One Year Later.” *Journal of Econometrics*.
- Cunha, Flávio, James Heckman, and Susanne Schennach.** 2010. “Estimating the Technology of Cognitive and Noncognitive Skill Formation.” NBER Working Paper 15664.
- Davis, Elizabeth E. and Aaron Sojourner.** 2021. “Increasing Federal Investment in Children’s Early Care and Education to Raise Quality, Access, and Affordability.” Washington, DC: Hamilton Project.
- Demir-Lira, Ö. Ece, Lauren R. Applebaum, Susan Goldin-Meadow, and Susan C. Levine.** 2019. “Parents’ Early Book Reading to Children: Relation to Children’s Later Language and Literacy Outcomes Controlling for Other Parent Language Input.” *Developmental Science* 22 (3): 1–16.
- Early, Diane M., Donna M. Bryant, Robert C. Pianta, Richard M. Clifford, Margaret R. Burchinal, Sharon Ritchie, Carollee Howes, and Oscar Barbarin.** 2006. “Are Teachers’ Education, Major, and Credentials Related to Classroom Quality and Children’s Academic Gains in Pre-Kindergarten?” *Early Childhood Research Quarterly* 21 (2): 174–95.
- Elango, Sneha, Jorge Luis García, James J. Heckman, and Andrés Hojman.** 2016. “Early Childhood Education.” In *Economics of Means-Tested Transfer Programs in the United States*, Vol. II, edited by Robert A. Moffitt, 235–97. Cambridge, MA: National Bureau of Economic Research.
- Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren and Michael Westberry.** 2021. “Integrated Public Use Microdata Series, Current Population Survey: Version 9.0 [dataset].” Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D030.V9.0>.
- Folbre, Nancy, Jayoung Yoon, Kade Finnoff, and Allison Sidle Fuligni.** 2005. “By What Measure? Family Time Devoted to Children in the United States.” *Demography* 42 (2): 373–90.
- García, Jorge Luis, James J. Heckman, and Victor Ronda.** 2021. “The Lasting Effects of Early Childhood Education on Promoting the Skills and Social Mobility of Disadvantaged African Americans.” NBER Working Paper 29057.
- Goerge, Robert, A. Rupa Datta, Kanru Xia, Ann D. Witte, Lisa A. Gennetian, Carolina Milesi, Richard Brandon, Lina Guzman, and Wladimir Zanoni.** 2015. *Identifying Head Start and Public Pre-K Participation in NSECE Data on Center-Based ECE Programs*. Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Guryan, Jonathan, Erik Hurst, and Melissa Kearney.** 2008. “Parental Education and Parental Time with Children.” *Journal of Economic Perspectives* 22 (3): 23–46.
- Heckman, James J., and Stefano Mosso.** 2014. “The Economics of Human Development and Social Mobility.” *Annual Review of Economics* 6 (1): 689–733.
- Heidlage, Jodi K., Jennifer E. Cunningham, Ann P. Kaiser, Carol M. Trivette, Erin E. Barton, Jennifer R. Frey, and Megan Y. Roberts.** 2020. “The Effects of Parent-Implemented Language Interventions on Child Linguistic Outcomes: A Meta-Analysis.” *Early Childhood Research Quarterly* 50 (1): 6–23.
- Hofferth, Sandra L., Sarah M. Flood, Matthew Sobek, and Daniel Backman.** 2020. “American Time Use Survey Data Extract Builder: Version 2.8 [dataset].” College Park, MD: University of Maryland and Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D060.V2.8>.
- Hotz, V. Joseph, and Matthew Wiswall.** 2019. “Child Care and Child Care Policy: Existing Policies, Their Effects, and Reforms.” *ANNALS of the American Academy of Political and Social Science* 686 (1): 310–38.
- Johnson, Anna D., Rebecca M. Ryan, and Jeanne Brooks-Gunn.** 2012. “Child-Care Subsidies: Do They Impact the Quality of Care Children Experience?” *Child Development* 83 (4): 1444–61.
- Larson, Anne L., Lauren M. Cychk, Judith J. Carta, Carol Scheffner Hammer, Melissa Baralt, Yuuko Uchikoshi, Zhe Gigi An, and Carla Wood.** 2020. “A Systematic Review of Language-Focused Interventions for Young Children from Culturally and Linguistically Diverse Backgrounds. *Early Childhood Research Quarterly* 50: 157–78.

- Maloney, Erin A. Benjamin A. Converse, Chloe R. Gibbs, Susan C. Levine, and Sian L. Beilock.** 2015. "Jump-Starting Early Childhood Education at Home: Early Learning, Parent Motivation, and Public Policy." *Perspectives on Psychological Science* 10 (6): 727–32.
- Mashburn, Andrew J., Robert C. Pianta, Bridget K. Hamre, Jason T. Downer, Oscar A. Barbarin, Donna Bryant, Margaret Burchinal, Diane M. Early, and Carollee Howes.** 2008. "Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills." *Child Development* 79 (3): 732–49.
- Mayer, Susan E., Ariel Kalil, Philip Oreopoulos, and Sebastian Gallegos.** 2019. "Using Behavioral Insights to Increase Parental Engagement: The Parents and Children Together Intervention." *Journal of Human Resources* 54 (4): 900–25.
- NORC.** 2012. "National Survey of Early Care and Education (NSECE) 2012." Ann Arbor: Inter-university Consortium for Political and Social Research. <https://doi.org/10.3886/ICPSR35519.v12> (accessed March 25, 2019).
- Pilarz, Alejandra Ros, Ying-Chun Lin, and Katherine A. Magnuson.** 2019. "Do Parental Work Hours and Nonstandard Schedules Explain Income-Based Gaps in Center-Based Early Care and Education Participation?" *Social Service Review* 93 (1): 55–95.
- Price, Joseph, and Ariel Kalil.** 2019. "The Effect of Mother–Child Reading Time on Children's Reading Skills: Evidence from Natural Within-Family Variation." *Child Development* 90 (6).
- Prickett, Kate C., and Jennifer March Augustine.** 2021. "Trends in Mothers' Parenting Time by Education and Work From 2003 to 2017." *Demography* 58 (3): 1065–91.
- Shonkoff, Jack P., and Deborah A. Phillips, eds.** 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academies Press.
- Survey Research Center.** 2019. "Panel Study of Income Dynamics: Public Use Dataset." Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.
- Todd, Petra E., and Kenneth I. Wolpin.** 2007. "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *Journal of Human Capital* 1 (1): 91–136.
- US Census Bureau.** 2018. "Survey of Income and Program Participation (SIPP)." 2018 Panel, Wave 1. Retrieved from <https://www.census.gov/programs-surveys/sipp/data/datasets.2018.html>.
- US Department of Education.** 2001–2006. "Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)." Washington, DC: National Center for Education Statistics.
- US Department of Education.** 2001–2019. "National Household Education Surveys Program (NHES)." Washington, DC: Institute of Education Sciences, National Center for Education Statistics. Retrieved from <https://nces.ed.gov/nhes/dataproducts.asp>.
- Watamura, Sarah Enos, Deborah A. Phillips, Taryn W. Morrissey, Kathleen McCartney, and Kristen Bub.** 2011. "Double Jeopardy: Poorer Social-Emotional Outcomes for Children in the NICHD SECCYD Experiencing Home and Child-Care Environments that Confer Risk." *Child Development* 82 (1): 48–65.
- York, Benjamin N., Susanna Loeb, and Christopher Doss.** 2019. "One Step at a Time: The Effects of an Early Literacy Text-Messaging Program for Parents of Preschoolers." *Journal of Human Resources* 54 (3): 537–66.

Economics of Foster Care

Anthony Bald, Joseph J. Doyle Jr., Max Gross, and
Brian A. Jacob

When authorities believe children have been abused or neglected, child protective services administered at the state or local level may remove those children from their homes and place them temporarily into substitute care. This practice, commonly known as foster care placement, involves a difficult tradeoff: protecting children versus preserving families. This tension has a long history in the US child welfare system. At the turn of the last century, “orphan trains” transported thousands of neglected children from cities to live and work on farms. Largely in response, early cash welfare programs, such as Mothers’ Pensions and Aid to Dependent Children, were implemented with the expressed intent of keeping children with their biological families (Aizer et al. 2016; Testa and Kelly 2020). The foster care system has changed over time, but the core debate over protecting children and preserving families remains. Today, foster care placement is common around the world: 5 percent of US children are placed in foster care at some point during childhood, and similar rates are found in other countries (Rouland and Vaithianathan 2018; Yi, Edwards, and Wildeman 2020).

■ *Anthony Bald is a PhD candidate in Public Policy, Harvard University, Cambridge, Massachusetts. Joseph J. Doyle Jr. is Erwin H. Schell Professor of Management and Applied Economics, MIT Sloan School of Management, Cambridge, Massachusetts. Max Gross is a Researcher at Mathematica, Princeton, New Jersey. Brian A. Jacob is Walter H. Annenberg Professor of Education Policy, Professor of Economics and Professor of Education at the Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, Michigan. Doyle and Jacob are both Research Associates at the National Bureau of Economic Research, Cambridge, Massachusetts. Their emails are abald@g.harvard.edu, jjdoyle@mit.edu, mgross@mathematica-mpr.com, and bajacob@umich.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.223>.

Children who spend time in foster care are particularly vulnerable. In comparison to similar children who did not spend time in foster care, foster children have rates of depression and anxiety that are seven times higher (Turney and Wildeman 2016) and also exhibit worse educational outcomes, such as lower test scores and graduation rates (Barrat and Berliner 2013). These disadvantages continue into adulthood. For example, close to 20 percent of the US prison population were in foster care as youth (US Bureau of Justice Statistics 2021). Those who turn 18 in foster care report 50 percent lower earnings and employment rates that are 20 percentage points lower by age 26 compared to a sample of young adults with similar levels of education (Okpych and Courtney 2014); as many as one-third experience homelessness (Dworsky, Napolitano, and Courtney 2013).

Economic theory can clarify the determinants of the demand and supply of foster care, including incentives to provide high-quality care. Econometric analysis can estimate the causal effects of foster care and related interventions on child well-being, and there is a growing body of empirical work on the topic. Federal and state policies now encourage the evaluation of foster care interventions and subsidize the implementation of evidence-based programs. These policies, along with increased data availability and an openness to testing new digital tools, have heightened the demand for rigorous research in the field.

In this paper, we provide an overview of the US foster care system and highlight key considerations in child welfare policy. We begin by discussing how the system works and show several prominent trends in foster care placement. We also highlight that states operate their child welfare systems quite differently, and the conditions states face and approaches they adopt have changed over time. Given this heterogeneity, we look within states and ask whether foster care placement improves outcomes for maltreated children.

We then turn to factors influencing the demand and supply of foster care. On the demand side, we discuss recent drivers of foster care, as well as prevention efforts through in-home family services and diversion from formal placement to less formal living arrangements with relatives. On the supply side, we describe the way the supervision of foster care is organized, efforts to increase the recruitment of foster homes, and the quality of care provided by different forms of foster care placement. We conclude by highlighting opportunities for future research.

The Landscape of Child Protection

Modern child welfare programs in the United States began with the establishment of the Children's Bureau in 1912, followed by federal funding in the Social Security Act in 1935. Indeed, Title IV of the Social Security Act is titled "Grants to States for Aid and Services to Needy Families with Children and for Child-Welfare Services." A series of federal and state laws addressing child maltreatment have been enacted since, and the competing goals of family preservation and child protection can be seen through the combination of federal funding of foster care placements

along with cash welfare and family unification initiatives (Haskins 2020). Child welfare agencies currently spend over \$30 billion each year on child protection, with approximately half of their funding provided by the federal government. Most recently, the Family First Prevention Services Act of 2019—commonly referred to as Family First—for the first time allows states to use federal funding on services designed to prevent foster care placement.

The path of a child into foster care most often starts with a phone call to report suspected child maltreatment. States typically classify child maltreatment into three categories: neglect, physical abuse, and sexual abuse. *Neglect* is the inability of caregivers to supply a child with adequate housing, food, clothing, or other basic needs; *physical abuse* is “any nonaccidental physical injury to the child”; and *sexual abuse* involves direct sexual contact or exploitation of children (Child Welfare Information Gateway 2019b). Roughly 54 percent of initial referrals are deemed to warrant further investigation. The remaining referrals are screened out because they do not fall under the state-specific definition of maltreatment, although such definitions often leave room for interpretation (Font and Maguire-Jack 2020).

People in certain occupations, such as educators, police officers, physicians, and social workers, are mandated by law to report suspected maltreatment to local authorities, though reporting requirements vary by state. The majority of screened-in reports come from mandated reporters. Educators make up 21 percent of reports, as they are frequently in close proximity with school-aged children and can observe signs of maltreatment. Law enforcement personnel are second at 19 percent, as they may request support from child protective services when responding to domestic disputes or reports of unsupervised children. Medical personnel and social workers are also common reporters (at 11 and 10 percent of reports respectively), as they may observe injuries that raise suspicion or signs of neglect in the home. The remainder of reports come from relatives, friends, and other miscellaneous sources (US Department of Health and Human Services 2021).

The Prevalence of Child Maltreatment and Foster Care Placement

In 2019, child protective services investigated nearly 3.5 million children for maltreatment (nearly 5 percent of all children), and classified 652,253 children as victims or nearly 1 percent of all children, as shown in Table 1.¹ Yearly incidence of maltreatment reported to child protective services dramatically understates the cumulative rate of maltreatment experienced by children. Over the course of childhood, a remarkable 37 percent of all children experience a child welfare

¹To characterize foster care practice across states and over time in this paper, we use data from the Adoption and Foster Care Analysis and Reporting System (AFCARS): <https://www.acf.hhs.gov/cb/data-research/adoption-fostercare/> (US Department of Health and Human Services 2020a). For child maltreatment statistics, we use the National Child Abuse and Neglect Data System (NCANDS), a centralized system for states to submit maltreatment reports to the federal government: <https://www.acf.hhs.gov/cb/research-data-technology/reporting-systems/ncands/>. In particular, we use NCANDS data that have been processed and combined with population data from the US Census by the Annie E. Casey Foundation (AECF): <https://datacenter.kidscount.org/> (Annie E. Casey Foundation 2022).

Table 1

Foster Care Trends

	<i>Number in</i> 2019 (1)	<i>Share in</i> 2004 (2)	<i>Share in</i> 2019 (3)	<i>Difference</i> (3) – (2) (4)
<i>Panel A. Children ages 0–17 (N = 73,039,150)</i>				
Investigated for maltreatment	3,449,674	3.52	4.72	1.19
Confirmed as victims	652,253	0.97	0.89	–0.08
Entered foster care	250,311	0.41	0.34	–0.06
In foster care at end of fiscal year	419,760	0.68	0.57	–0.11
<i>Panel B. Removal reason for children entering care (N = 250,311)</i>				
Removed due to neglect		51.43	63.87	12.44
Removed due to parent substance use		23.34	38.15	14.81
Removed due to physical abuse		16.88	12.94	–3.94
Removed due to sexual abuse		6.33	3.96	–2.37
<i>Panel C. Placement setting for children at end of year (N = 418,654)</i>				
Unrelated foster family		47.40	46.62	–0.77
Kinship care		22.61	31.94	9.33
Congregate care		18.05	10.56	–7.49
Other setting		11.95	10.88	–1.07
<i>Panel D. Exit reason for children leaving care (N = 241,796)</i>				
Reunification with parents		54.20	48.32	–5.88
Adoption		18.90	26.63	7.72
Transition to relatives or guardianship		16.91	17.20	0.29
Emancipation		5.95	6.40	0.45

Notes: Reported sample sizes are for 2019. Panel A reports statistics among children ages 0–17. Panel B reports the percentage of children removed for each reason among children entering foster care. Children can be removed for multiple reasons, which are not mutually exclusive. Panel C reports the most recently observed placement setting for children in care at the end of the fiscal year. This sample size is smaller than the total reported in panel A due to missing data on placement setting. Panel D reports exit reasons for children leaving care. Exits from care are mutually exclusive, where the omitted category (1.45 percent) includes agency transfers, runaways, and child deaths. See text for sources.

investigation, and 12 percent of children are identified as maltreatment victims, including 18 percent of Black children and 16 percent of Native American children (Kim et al. 2017; Yi, Edwards, and Wildeman 2020).² These numbers only capture child maltreatment reported to authorities; self-reported victimization rates in survey data reach as high as 38 percent (Finkelhor et al. 2015).

Children identified as maltreatment victims typically remain with their family. Many of these children’s families are provided with services, such as family therapy,

²While the lack of comparable data makes it difficult to compare child maltreatment across countries, some evidence suggests that the United States ranks toward the top of more objective indicators of child maltreatment, such as child fatality due to intentional injury (OECD 2013, Chart PF1.9.A).

parenting programs, or substance abuse treatment. The remaining maltreatment victims are removed from their families and placed in foster care.

Table 1 shows that there were over 400,000 children in foster care at the end of 2019, or 0.6 percent of the child population—down from 0.7 percent in 2004. The share of children in care has fluctuated over time, rising from 300,000 in the 1980s to over 550,000 by the early 2000s (Testa and Kelly 2020). Trends in child maltreatment and foster care placement were somewhat more stable from 2004–2019. Panel B of Table 1 shows that the allegations associated with child removals have shifted substantially toward neglect (64 percent of children in 2019) and away from abuse (17 percent of children in 2019). Moreover, Wulczyn (2020) documents demographic shifts in the types of children who are entering foster care, with greater proportions entering as infants and smaller shares from large urban centers since 2000. As we explain discuss below, there are several potential explanations for this shift, including the opioid epidemic.

Placement Types and Time Spent in Foster Care

Children who are removed from home can be placed in several different types of foster care. Placement with an unrelated foster family is most common, comprising about half of children in foster care at the end of 2019, as shown in Panel C of Table 1. One of the major developments in foster care over the past few decades is the increased use of kinship care, a practice whereby children are placed with a nearby relative, such as a grandparent. In Table 1, the share of foster children in kinship care was 32 percent at the end of 2019, an increase of 9 percentage points since 2004. Just over 10 percent of foster children were placed in congregate care settings such as group homes and institutions, a rate that typically varies from 5 to 20 percent across states. As evidence has mounted showing poor outcomes for children placed in congregate care, its use has fallen out of favor (Lee et al. 2011). For instance, under the 2019 Family First legislation, federal funding only covers the first two weeks of placement into congregate care.

The typical goal for a foster child is a return home, although for some the goal is placement in an adoptive home. In 2019, 48 percent of all children who exited foster care were reunified with their families, a 6 percentage-point decrease from 2004 (Panel D of Table 1). The next most common exit types are adoption (27 percent), living with relatives or guardians (17 percent), and emancipation as a legal adult (6 percent).³ Adoption in particular has become more common in recent years, following legislation in the 1990s to incentivize shorter stays (Haskins 2020).

In terms of the foster care experience, among children who entered foster care in 2015, the median length of stay was 15 months, although some children stay much longer. Ten percent of children entering in 2015 were still in foster care four

³The AFCARS data used in Table 1 is not the ideal data set for measuring high frequency entry and exit dynamics. The underlying state-level data is collected in semi-annual reporting windows, and states report information from a child's most recent removal event in a given window.

years later, but in three states this fraction exceeded 20 percent. With longer stays often come multiple foster homes, and placement stability is often considered a quality metric for child welfare systems due to its strong correlation with behavioral outcomes (Rubin et al. 2004). At the end of the 2019 fiscal year, 35 percent of children in foster care had lived in three or more different settings.

Racial Disparities in the Child Welfare System

Racial disparities exist at every decision point in the child welfare system, resulting in large differences in involvement with child protective services by the time children reach age 18. For instance, Black children are reported to child protective services more frequently than White children, and reports involving Black children are more likely to be confirmed as maltreatment (Putnam-Hornstein et al. 2013). Furthermore, an estimated 10–12 percent of Black children and 11–15 percent of American Indian/Alaska Native children will experience a foster care placement over the course of their childhood, compared to 5 percent of White children (Wildeman and Emanuel 2014). Understanding the causes and consequences of disparities, and what policies can be enacted to reduce them, is an active area of research (for overviews, see Barth et al. 2020; Dettlaff and Boyd 2020).

Heterogeneity across Child Welfare Systems

Child welfare systems are administered at the state or local level; therefore the emphasis on child protection versus family preservation can fluctuate over time and across jurisdictions. For example, there are prominent examples of a community experiencing a tragedy, such as a case of severe abuse or a child fatality, that led to policy changes emphasizing child protection followed by a steep rise in placement rates. Years later, the higher cost of maintaining a larger share of children in foster care may be met with reforms to speed reunification and reduce placement rates through family preservation programs.⁴

Partially as a result of local policies, foster care placement rates vary substantially across states. Figure 1a shows the distribution of foster care placement rates across states in 2019, which range from under 3 per 1,000 children in New Jersey, Utah, and Virginia to over 15 per 1,000 children in West Virginia, Montana, and Alaska. Even within states, there have been considerable changes over time. Between 2004 and 2019, the rate per 1,000 children in foster care nearly doubled in West Virginia (10 to 20.2) and Montana (9.1 to 16.1), but fell substantially in Nebraska (14 to 7.1) and the District of Columbia (21.6 to 5.2). Such heterogeneity may also reflect demand-side factors, although they are not well explained by differences in official rates of maltreatment.⁵

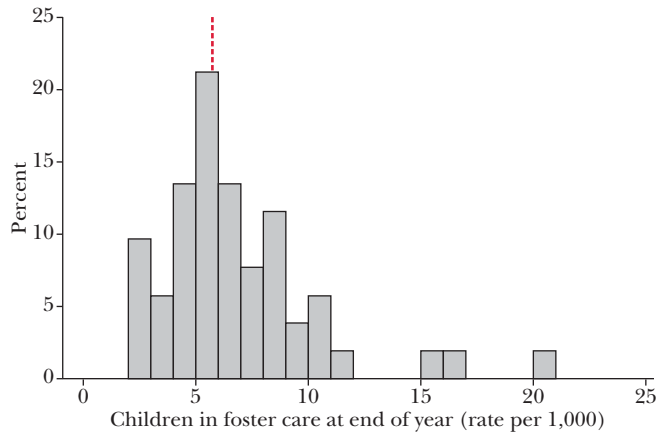
⁴The State Child Abuse & Neglect Policies Database (<https://www.scanpoliciesdatabase.com/>), which provides detailed state-level information on maltreatment definition and policy context, is a useful resource to understand the varying policy landscapes.

⁵Reported maltreatment rates explain about 15 to 28 percent of the variation in the percentage of children entering foster care across states. In a state-level regression of the percentage of children entering foster care in 2019 on the percentage of children subject to an investigated maltreatment report in 2019,

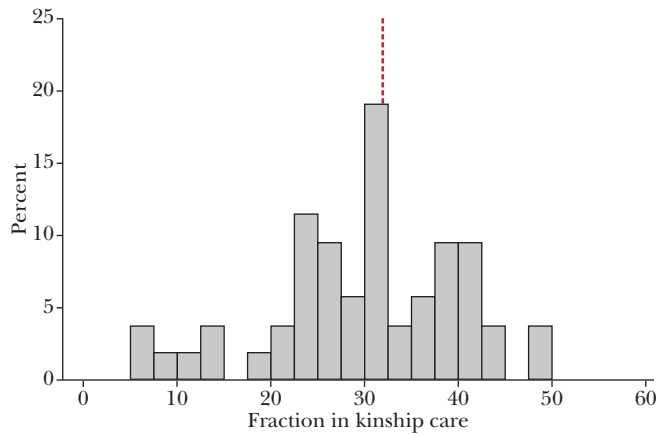
Figure 1

State Variation in Foster Care Placement Rate and Percentage in Kinship Care

A: Histogram of children in foster care at end of year (rate per 1,000)



B: Histogram of percentage of children in kinship care



Source: See text.

Notes: Panel A reports the foster care placement rate per 1,000 children (one observation for each state). We report placement rates among children ages 0 to 17 who were in foster care at the end of the 2019 fiscal year. The vertical line represents the nationwide placement rate. Panel B reports the percentage of foster children in kinship care (one observation for each state). We report the percentage of children ages 0 to 17 who were in foster care at the end of the 2019 fiscal year and who were placed in kinship care. The vertical line represents the nationwide percentage in kinship care.

Similarly, while kinship care typically receives priority when placements are considered, states vary widely in their reliance on kin. Figure 1b shows kin placement

the R-squared is 0.28. Using the percentage of children confirmed as victims of maltreatment gives an R-squared of 0.15.

rates range from less than 10 percent to nearly 50 percent across states. This variation in kinship placement stems in large part from differences in policy emphasis of recruiting and certifying kin, which mirrors the ongoing tradeoff between child protection and family preservation. Given the variation in child removal rates and placement types across child welfare systems, more evidence on the types of interventions that promote child well-being is needed.

For more detailed information on these trends and heterogeneity across states, the Appendix provides a series of figures. These include trends in child maltreatment, foster care placement, reasons for placement, and placement types over time and across states.

Family Preservation versus Child Protection

When deciding on the stringency of child welfare policy, a central question is whether foster care placement improves child welfare *on the margin*. In other words, what is the impact of foster care placement for a child who was removed under one regime, but would *not* have been removed under an alternative regime? Although a well-documented correlation exists between foster care and poor outcomes for children even after controlling for observable differences across cases (for example, Berger et al. 2015), it is challenging to identify causal effects because foster children differ in unobserved ways from children who were not placed. For example, foster children may have been more severely abused or neglected.

In the economics literature, the main empirical approach to overcome these confounding factors has been to consider the quasi-random assignment of child protective investigators to cases. When a report of maltreatment is made to a local child welfare office, a case worker is sent to investigate the report and determine if there is sufficient evidence of abuse or neglect. Many localities use a rotation process to assign cases to investigators working in the same field office to smooth the workload and often out of concern for fairness. In addition, despite assignment to similar cases, some investigators are more likely to recommend foster care placement compared to their colleagues. It is therefore possible that two children with very similar backgrounds will face different likelihoods of placement by virtue of receiving different (as-if randomly assigned) investigators. Such comparisons offered by this approach recover the causal effects of foster care for *marginal* children: those for whom the investigator assignment matters for the foster care placement decision. As a result, an important limitation of this approach is that it cannot determine effects for cases when all investigators would remove the child, nor for cases when no investigator would remove the child. Even so, understanding outcomes for marginal children provides insight into the effects of more- or less-stringent child welfare policies.

Doyle (2007a) used this approach to study children placed into foster care in Illinois during the 1990s. The study linked maltreatment and foster care data to state administrative data to measure a wide range of longer-term outcomes. The

sample included children who were receiving welfare benefits (which facilitated matching across the outcomes) and who were aged 5–15 at the time of the initial investigation. The results imply large negative effects of foster care placement on the margin, including substantially lower earnings and large increases in teen motherhood, delinquency, and unemployment. In later work, Doyle (2008, 2013) shows that foster care placement also increased the likelihood of both emergency medical care within a year of placement and criminal justice system involvement in adulthood. In similar work using the investigator strategy in South Carolina, Roberts (2019) finds that foster care placement roughly doubles the likelihood of juvenile delinquency over the five years following an investigation (an 11 percentage-point increase).

In contrast, positive effects have been found in other contexts. Gross and Baron (forthcoming) studied school-aged children in Michigan subject to investigations from 2008 to 2016. The study finds that removal causes a significant reduction in the likelihood of subsequent maltreatment, as well as improvements in math test scores and school attendance. In related work, Baron and Gross (2022) find that removal significantly reduces adult criminality. In South Carolina, Roberts (2019) finds that foster care placement reduces the likelihood of repeating a grade by 13 percent within three years.

In addition, there is evidence that the effects of foster care placement vary across child subgroups, with younger children tending to benefit more. Bald et al. (forthcoming) used linked administrative data from Rhode Island and found a positive impact of removal on measures of school achievement concentrated among girls removed before age six (they do not find such impacts for boys). Gross and Baron (forthcoming) also find improved educational outcomes for younger children, while others find worse outcomes for older children in multiple settings (Doyle 2007a, 2008; Warburton et al. 2014).

A possible reason for the contrasting findings across states and time periods is their different institutional contexts, as effects of foster care may vary with the threshold for placement and the quality of the foster care system. For example, foster care placements in Illinois in the 1990s were substantially longer and less stable than recent placements in Michigan.

Within a given context, studies have estimated the effects of foster care at different margins determined by different types of investigators. For example, among a subset of investigators with relatively low removal rates (more “lenient” investigators), the marginal child likely has relatively more severe maltreatment characteristics, as these investigators have a higher threshold before they recommend placement. Doyle (2007a) finds that the negative effects of foster care in terms of teen motherhood are larger among children assigned to investigators with low removal rates and, consequently, marginal cases that have relatively severe maltreatment compared to cases investigated by higher-removal-rate investigators. Meanwhile, negative effects on labor market outcomes are found to be similar across investigator types, suggesting that the placement decision resulted in worse employment outcomes across the severity spectrum. In contrast to Doyle (2007a),

Bald et al. (forthcoming) find evidence of effects on achievement that are especially positive for young girls with higher unobserved maltreatment. This suggests that young girls with high risk levels experienced the highest gains.

One potential mechanism for the way involvement in child protective services affects children is how such involvement affects their family members, some of whom may be identified as “perpetrators” of maltreatment. The child protective services authorities provide a variety of services to families with the goal of preventing the need for removal, or in cases where removal is unavoidable, to support family reunification and decrease risk levels (Merritt 2020). Gross and Baron (forthcoming) find that in Michigan, children’s gains from foster care placement occurred after most were reunified with their birth parents, which is consistent with the rehabilitation of birth parents while their children were in foster care. Similarly, using data from Allegheny County, Pennsylvania, Grimon (2021) finds that mothers were substantially more likely to enter drug treatment and receive mental health services not only at the time of the investigation, but for many years afterward.

Given the mixed evidence on the effects of placement for different margins and contexts, understanding the types of cases and child welfare interventions that promote child wellbeing is an important area for future research. One way to characterize the sets of issues involved is to focus on the demand and supply of foster care services.

Foster Care Demand

Demand for foster care comes from both the extent of child maltreatment in an area and policies that dictate how child protective services responds to it. As shown in Table 1, victimization rates have been relatively steady over the past 20 years, but the nature of maltreatment has moved away from child abuse toward child neglect. Partly as a result, the ways in which states respond has changed over time as well, tending toward policies that aim to prevent the need for foster care.

Risk Factors

Poverty is strongly associated with child maltreatment (for a survey of the literature, see Bullinger, Lindo, and Schaller forthcoming). For example, numerous studies have shown that foreclosure, eviction, and other forms of housing insecurity, as well as food insecurity, increase reported child maltreatment. Less directly, financial resources can also influence parental stress, which could in turn lead to abuse or neglect.

Given these associations, efforts to reduce poverty, or to ameliorate its effects, can reduce demand for foster care. Experimental studies of welfare reforms suggest a causal relationship between family income and child maltreatment (Fein and Lee 2003; Cancian, Yang, and Slack 2013). These studies sometimes capture multiple policies (for example, changes in benefit levels as well as work

incentives) and might not extend to different contexts. Quasi-experimental analyses of earned income tax credits across states find higher benefit levels are associated with a reduction in maltreatment rates, particularly for child neglect (Berger et al. 2017). Other studies find lower rates of maltreatment associated with investments in the social safety net.

Another leading driver of neglect is parental substance abuse, which has increased substantially over time as a fraction of maltreatment reports (see Table 1 earlier). Studies show that the ongoing opioid epidemic in particular has contributed to child maltreatment and foster care. Although it is challenging to measure the effects of the opioid epidemic, studies have leveraged changes in the features of the crisis, such as shocks to illicit drug prices, treatment, or opioid supply, and conclude that the opioid epidemic has increased foster care caseloads (for example, Hou 2021). Taken together, the evidence indicates that policies outside of child welfare can influence the extent of child maltreatment and the demand for foster care.

The ways in which authorities from child protective services evaluate risk factors can influence demand as well. Historically, child protective services has used relatively rudimentary risk-modeling software to help caseworkers assess the potential harm a child faces by remaining at home. More recently, data analytics are showing promise in improving risk stratification, although concerns about accuracy, equity, and fairness are prominent. In Allegheny County, Pennsylvania, the use of more sophisticated software to predict the likelihood of future out-of-home placement has been incorporated into the workflow of screening-in alleged maltreatment for further review. Efforts in Douglas County, Colorado, to display this kind of risk information to investigators are being studied with randomized controlled trials (Vaithianathan 2019). As with other diagnostic tools, equity concerns are important to consider. In Allegheny, the introduction of decision support was associated with a reduction in the Black-White gap in screen-in rates—primarily as a result of increasing the percentage of White children that were referred for investigation (Goldhaber-Fiebert and Prince 2019). Research will be needed to determine not only the efficacy of these tools, but also the potential existence of biases built into them (Drake et al. 2020).⁶

Family Preservation

Child welfare agencies employ various strategies to reduce the demand for foster care by stabilizing at-risk families. The 2019 Family First legislation allows states to accept federal funds that were previously earmarked for foster care services and use them for prevention services. Over time, the goal is that such funding will become restricted to prevention services that have empirical support in one

⁶There are ongoing efforts to blind child welfare investigations to race by removing racial indicators and correlates from investigation reports (Baron, Goldstein, and Ryan 2021). But some caution is warranted in taking this step, because blinding algorithms to indicators of race may increase inequity (Rambachan et al. 2020).

of three areas: in-home parent skill-based programs, mental health, and substance abuse. The Title IV-E Prevention Services Clearinghouse contains a list of approved programs (available at <https://preventionservices.abtsites.com/>). States also apply for federal waivers to experiment with new approaches (James Bell Associates 2019).

Family First calls for each state to submit a five-year plan with proposed interventions for funding eligibility, which provides a glimpse of the types of programs that are popular to prevent foster care placement. For example, a common in-home parenting skills intervention is Parents as Teachers, an education program that aims to increase school readiness and reduce child maltreatment. The Nurse Family Partnership, a home visiting program that educates and supports new mothers, is common in part because experimental evidence shows the program results in a reduction of child neglect (Olds et al. 1997). For the other two categories of interventions, mental health and substance use, there are several trauma-informed, therapy-based interventions available. A common program is Functional Family Therapy, which includes weekly therapy sessions to build skills in youth and in the family as a whole.

A possible concern is that federal funding for family preservation programs could increase the surveillance of families. Monitoring families whose children remain in the home can induce fear that their children will be removed (Fong 2020). Increased surveillance may also create challenges for studies that use maltreatment reports as an outcome by artificially increasing reports. As a result, evaluations of prevention programs should focus on a wide range measures of well-being, including health, education, and criminal justice involvement outcomes.

A more controversial approach to preventing foster care is known as diversion, in which child protective services encourages the informal placement of children with substitute caregivers, often kin, rather than entering the formal foster care system (Berrick and Hernandez 2016). Opponents argue that diversion limits a family's access to services because non-licensed relatives do not qualify for the same monthly payments or family rehabilitation services as licensed foster parents. Advocates of diversion claim that the practice helps to keep families intact by allowing relatives, who may not meet restrictive licensing requirements, to care for children. A 2007 survey found that 39 states engaged in diversion (29 states actively promoted the practice), while 12 states prohibited it (Annie E. Casey Foundation 2013). Recently, Family First provided funding for Kinship Navigator programs, which vary in implementation but aim to provide a middle ground between diversion and formal placement.

Understanding the effect of diversion can help inform the tradeoff between family preservation and child protection. Unfortunately, scant evidence exists on the effects of diversion. Chen et al. (2020) evaluated Safe Families for Children, a program that placed children with foster families supervised by a nonprofit agency, but without formal placement in the foster care system. Using a randomized encouragement design across cases, they found this program increased the share of children who had been returned home at one year, while not leading to greater subsequent reports of abuse.

Extending Care to Older Ages

Demand for foster care can increase by extending the age of eligibility. Some 20,000 foster youth aged out of foster care in 2019 (based on the AFCARS data), and such youth may have few lifelines as they enter their 20s. Courtney et al. (2020) find that, among “emancipated” former foster youth in California, one in five previously dropped out of high school, one in four experienced homelessness, and nearly a third were incarcerated, all by the age of 23. One estimate suggests that the United States would save over \$4 billion in lifetime costs if these youth had outcomes comparable to the general population (Annie E. Casey Foundation 2019). The Fostering Connections Act of 2008 aimed to provide additional support for these youth by permitting states to use federal funds for foster youth beyond age 18.

A small body of descriptive and quasi-experimental research suggests that the push for extended eligibility provided a wide range of benefits to affected youth. Using nationwide data, Prettyman (2020) leverages the staggered roll-out of extended foster care across states. Using variation within states over time, the study finds that extended foster care reduces the likelihood of homelessness and incarceration, and increases high school graduation and employment. The California Youth Transitions to Adulthood Study (CalYOUTH) shows similar improvements in outcomes (Courtney, Okpych, and Park 2018). Increasing demand for foster care services by expanding the eligible population to include vulnerable youth over the age of 18 appears to have been a wide-ranging policy success.

Foster Care Supply

The impacts of foster care also depend on the functioning of the child welfare system on the supply side. Considerations include the public or private supervision of foster care, ways of recruiting, licensing, and assigning foster homes, and reliance on different types of foster care.

Public and Private Supervision

Once children are placed in foster care, they are supervised by a separate set of caseworkers who specialize in recruiting and monitoring foster homes, reunifying families, and finding adoptive homes. These tasks are often provided by private organizations (such as religious organizations). In the last few decades, several states have privatized some or all of their child welfare services. States award contracts to private non-profits and for-profit firms to assume responsibilities such as monitoring the safety of children in care.

Privatization can come with potential challenges, including stakeholder buy-in and accurate cost estimates (US Department of Health and Human Services 2007). Hubel et al. (2013) studied Nebraska’s effort to privatize its child welfare system, which involved replacing the traditional fee-for-service model with a capitation amount based on an expected cost of service

delivery.⁷ The authors find that initial payment rates were too low, leading to substantial problems—namely, reductions in quality and availability of services along with a 27 percent increase in costs.

Performance-based contracts in state foster care systems show some promise, however. Such initiatives often include incentive payments (for instance, linking some portion of provider payment or future placements to measurable child outcomes). In addition, agency officials and providers review performance data, discuss client outcomes, and attempt to identify and institutionalize effective strategies for serving clients. Reilly et al. (2021) find that 25 states currently use performance-based contracting for at least one child welfare service. Several decades ago in New York, a randomized evaluation of a family reunification program, administered by private agencies on a capitation basis, found mixed results (Westat et al. 1998).

Although we are not aware of any recent experimental evaluations of performance-based contracting in the field of foster care, descriptive studies suggest that such initiatives may be effective. Garstka et al. (2012) report on demonstration projects that implemented performance-based contracting in Florida, Missouri, and Illinois between 2007 and 2010. They identify several metrics that improved after implementation, including increased child safety and successful discharge from care. Similarly, case studies of Tennessee and Rhode Island describe their experiences with performance-based contracting as successful (Lawler and Foster 2013; Government Performance Lab 2019).

Regardless of the payment model, a better understanding of heterogeneity in value-added created by foster care agencies would inform improvements to the system. Quality dimensions at the state level are considered by federal Children and Family Services Reviews, which provide a blueprint for outcomes that might be considered. These include measures of safety (preventing foster care reentry), permanency (returning children home or providing placement stability), and child and family wellbeing. The wellbeing measures consider the receipt of needed services, although a wider set of outcomes could be included, such as educational outcomes and criminal justice involvement.

Foster Home Recruitment

Perhaps the simplest way to affect the supply of foster homes is to recruit more foster homes. Having a large number of willing foster homes can also result in better matching of foster children to homes (Hansen and Hansen 2006). Modeling match quality using administrative data from Los Angeles, Robinson-Cortes (2019) shows that enlarging the pool of foster homes available to children would likely improve child outcomes. Recruiting more foster homes may also

⁷Fee-for-service models in child welfare are structured so that the provider receives payments based on the specific services provided to foster children. A capitation model, in contrast, uses cost estimates to provide a fixed payment to the provider per child served, shifting more of the financial risk onto providers.

reduce the need for congregate care, which is seen as a last resort for children who cannot or should not be placed in a family setting (Chadwick Center and Chapin Hall 2016).

The federal Children's Bureau recently assessed states on their foster home recruitment practices and rated 34 states as needing improvement (US Department of Health and Human Services 2020b). While no national dataset exists to measure foster home capacity, the amount of foster homes available in some states outnumbers foster placements (Wulczyn et al. 2018). This suggests that there is not a shortage of foster homes everywhere, although child welfare agencies note that it can be difficult to find homes for subgroups like older children and children with special needs (Annie E. Casey Foundation 2018). Agencies often rely on recruitment events at places of worship and advertising in partnership with other private organizations. Recently, digital and data-driven tools have been adopted to support recruitment efforts.⁸ A better understanding of the methods available to modernize foster home recruitment may yield substantial gains in match quality.

Foster Home Licensure

A key issue that affects the supply of foster homes is the set of requirements to obtain a foster home license. All states require prospective foster families to undergo criminal background checks, but important details of licensing vary across states. For instance, Table 2 shows that as of 2018, nine states disqualify applicants for any drug-related crime, 14 require citizenship, 18 have minimum square footage requirements, 18 require training in first aid, and 26 require a non-smoking environment. Further, foster parents undergo 23 hours of initial training on average, which varies substantially across states (the standard deviation of initial required training is 8 hours). They also complete 12 hours of annual retraining (with a standard deviation of 4 hours).⁹ For kinship placements, relatives can be unlicensed in some states as long as basic safety standards, such as passing background checks, are met.

States do not receive federal subsidies for unlicensed homes, so they have incentives to encourage licensure. They may also have an incentive to lower licensure requirements, so the US Administration for Children and Families recently released a set of minimum quality standards for licensing foster homes, which includes six hours of pre-licensure parenting training, a written evacuation plan in case of an emergency, and assurance that the foster parents will not smoke in the presence of the child. Although the recommended standards are not binding, states are required to submit plans with a discussion of how and why their licensing standards deviate from the national standards (Children's Defense Fund 2020).

⁸For example, see a description of New York's recruitment efforts at <https://ocfs.ny.gov/programs/fostercare/recruitment/>.

⁹Authors' calculations (see Table 2 notes). Statistics on references, initial training hours, and annual training hours could be calculated for 44 states, 42 states, and 42 states, respectively. The remainder of statistics in Table 2 are derived from Child Welfare Information Gateway (2018, 2019a).

Table 2
State Foster Home Licensing Requirements

	<i>States</i>
<i>Criminal history requirements:</i>	
Background check and child abuse registry check	50
Child abuse registry check for previous states lived in	35
Never convicted of assault or battery	13
Never convicted of a drug-related crime	9
Juvenile court records check	7
<i>Individual requirements:</i>	
Aged 21 or older	36
US citizen or legal resident	14
Legally married if a couple	5
Ability to communicate in English	4
<i>Training requirements:</i>	
Initial licensing training	45
Annual training for license renewal	42
Training in First Aid/CPR	18
<i>Home requirements:</i>	
Separate bedrooms for opposite-sex children	37
Secure firearm storage	35
Non-smoking environment	26
Carbon monoxide detectors	20
Minimum square footage per child	18

Notes: State requirements come from the Child Welfare Information Gateway (CWIG 2018, 2019a). Criminal history requirements are current as of September 2018. Individual, training, and home requirements are current as of February 2018. Note that some state child welfare systems are managed at the county level and may have stricter or more lenient requirements.

It is not clear from economic theory how licensing requirements affect the well-being of foster children. On one hand, they establish minimum quality standards for specific aspects of care. On the other hand, licensing requirements may limit the supply of foster homes, which could increase placement in congregate care settings and reduce the quality of matches among those placed with families. In part because of the many requirements, it can take months for homes to obtain a license. Digital aids to complete and submit paperwork should increase the efficiency of the process. Regardless, further research on the effects of different licensing requirements is needed.

Monthly Maintenance Payments

Financial incentives can also affect the number of families willing to take in foster children. Foster families receive monthly subsidies to support the children staying with them, as well as one-time payments for specific needs. The average

monthly foster care payment was just over \$500, which is twice the amount of the average Temporary Assistance for Needy Families benefit (US GAO 2011). Payments vary widely across states, as well as by the age of the child, relative status, and other factors.

In theory, subsidies could act as a price mechanism to clear the market, but in practice, they reflect rules of thumb tied to the average cost of hosting one child. For instance, foster care subsidies often do not vary by location within a state or by placement setting, are not adjusted for inflation, and are revised infrequently (Goldhaber-Fiebert et al. 2014).

Some evidence suggests that these subsidies could be used as a policy lever. Doyle and Peters (2007) studied foster care supply and demand in the early 1990s. The study estimates that the monthly subsidy has a supply elasticity of 0.3 in states with low initial subsidy levels. In a related paper, Doyle (2007b) finds that when Illinois reduced the monthly subsidy to relatives by 30 percent, relatives were 15 percent less likely to offer care. The finding that foster homes respond to financial incentives is echoed in work on adoption subsidies, which shows that raising adoption subsidies increases adoptions relative to long-term foster care (for example, see Brehm 2021 and the references therein).

Effects of Different Types of Placements

The type of placements that are employed is another supply-side policy that can affect child wellbeing. Foster care placements can vary in a number of ways: for example, kin or non-kin families, or residential versus family care. Non-kin families typically receive more training and vetting, and they are volunteering their services, which may lead to higher quality care. Meanwhile, kinship foster parents are typically offered fewer supports from child welfare agencies (Sakai, Lin, and Flores 2011), though kinship care may benefit children by retaining family connections. Similarly, residential care may provide specialized mental or physical health services.

Evidence on the effects of foster placement type is limited, but some quasi-experimental studies shed light on the relative benefits of placement types. Doyle's (2007b) examination of the reduction of monthly payments made to kin found that measures of quality of care, including wellness visits and placement stability, were not affected. This suggests that marginal kinship providers are similar to non-kin providers in terms of these quality measures. Hayduk (2017) studied state policy changes and found that children exposed to laws favoring kinship care experience more stable placements and shorter foster care episodes, but mental and physical health outcomes appear unaffected in the short run.

Independent Living Programs

When children transition out of foster homes or congregate care as young adults, independent living programs aim to help prepare them for adulthood. All youth ages 14 to 23 currently or formerly in foster care are eligible for independent living programs (Children's Defense Fund 2020). Such programs include assistance in completing a high school diploma or GED, job training, or postsecondary

education. Young adults exiting the foster care system who are at risk of homelessness are also eligible for housing vouchers provided by the US Department of Housing and Urban Development.

The federal government funded four experimental evaluations as part of the creation of the Chafee Foster Care Independence program enacted in 1999. Just one out of the four studies found positive impacts on key indicators of a successful transition to adulthood. The lone success story was from the Massachusetts Adolescent Outreach Program for Youths in Intensive Foster Care, which paired youths with a caseworker who met with young people once a week for 16 months to offer individualized services in areas such as job search. The program improved college enrollment and persistence but did not have impacts on other outcomes that it sought to influence, like employment and housing (Courtney et al. 2011). More recent evaluations of independent living programs show modest promise (for example, Skemer and Valentine 2016).

Conclusion

Foster care is a far-reaching intervention in the lives of particularly vulnerable children. With substantial variation in how child welfare systems use the practice, and emerging evidence of heterogeneous effects across systems, more evidence is needed to inform policy challenges and improve child and family wellbeing.

On the demand side, a better understanding of the drivers of child maltreatment, and interventions that ameliorate them, could prevent the need for foster care in the first place. This suggests that poverty alleviation programs, as well as programs aimed to improve the human capital of parents, should consider child maltreatment and foster care placement as outcomes in their evaluations. Further, targeted family preservation initiatives now receive federal subsidies, and there is an appetite for more evidence on the effectiveness of such programs. There is also a need for better evidence on diversion programs that circumvent the formal foster care system, which are not only controversial but also understudied.

On the supply side, efforts to improve the quality of foster care through modernizing the recruitment and matching of foster homes are gaining some popularity and are in need of rigorous evaluation. Better measures of the supply of foster homes and signs of shortages, such as placement distance-from-home and time in emergency shelters, would help to evaluate the extent of any shortage problem and to gauge the effectiveness of efforts to solve it. For example, returns to different forms of targeted advertising, including appeals to child wellbeing, monthly subsidies, or information about the licensure process should be estimated to inform future recruitment strategies.

Some key questions about the best type of placement—non-kin, kin, and congregate care—for different types of children remain unanswered. For example, children currently residing in congregate care could have their placements revisited through more-intense case management, to test whether time in congregate

care can be reduced and whether subsequent outcomes improve. In addition, the services provided to children already in foster care deserve more attention. Counseling, mentorship initiatives, and jobs programs should all be studied. At the agency level, much more could be learned about the impacts of private provision of services and performance contracting.

As agencies work to advance equity in child welfare and foster care, studying the effects of programs and policies are crucial next steps. For example, several jurisdictions have considered removing racial indicators and correlates from investigation reports when determining whether to place a child in foster care, yet research suggests caution is warranted in taking this step (Baron, Goldstein, and Ryan 2021).

More generally, understanding the mechanisms driving heterogeneous effects of foster care is an important next step in addressing the underlying tradeoff between child protection and family preservation. The current patchwork of state and local foster care systems provides an opportunity for well-identified studies to understand the benefits and costs of such systems. Meanwhile, linked administrative data and foster care program data, after two decades of federal subsidies to modernize state child welfare data systems, are providing new opportunities to study the long-term impact of interventions. Coupled with a willingness among policy-makers to generate and employ evidence on best practices, foster care research is poised to inform policy and improve the welfare of children.

■ *We are grateful to Lawrence Berger, Eric Chyn, Kim Dadisman, Erik Hurst, Barrett Johnson, Bridgette Lery, Margarita Machelett, Enrico Moretti, Nina Pavcnik, Heidi Williams, and Fred Wulczyn for helpful comments, and especially to Timothy Taylor for extensive comments and edits. We thank Sam Haas for excellent research assistance. The data used in this publication, AFCARS Foster Care Files (1995–2019), were obtained from the National Data Archive on Child Abuse and Neglect and have been used in accordance with its Terms of Use Agreement license. The Administration on Children, Youth, and Families, the Children’s Bureau, the original data set collection personnel or funding source, NDACAN, Cornell University and their agents or employees bear no responsibility for the analyses or interpretations presented here.*

References

- Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. “The Long-Run Impact of Cash Transfers to Poor Families.” *American Economic Review* 106 (4): 935–71.
- Annie E. Casey Foundation. 2013. *The Kinship Diversion Debate: Policy and Practice Implications for Children, Families and Child Welfare Agencies*. Baltimore: Annie E. Casey Foundation.

- Annie E. Casey Foundation.** 2018. *What Impacts Placement Stability?* Baltimore: Annie E. Casey Foundation.
- Annie E. Casey Foundation.** 2019. *Future Savings: The Economic Potential of Successful Transitions from Foster Care to Adulthood.* Baltimore: Annie E. Casey Foundation.
- Annie E. Casey Foundation.** 2022. KIDS COUNT Data Center. Annie E. Casey Foundation. <http://datacenter.kidscount.org/>.
- Bald, Anthony, Eric Chyn, Justine S. Hastings, and Margarita Machelett.** Forthcoming. "The Causal Impact of Removing Children from Abusive and Neglectful Homes." *Journal of Political Economy*.
- Bald, Anthony, Joseph J. Doyle Jr., Max Gross, and Brian A. Jacob.** 2022. "Data and Code for: Economics of Foster Care." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <http://doi.org/10.3886/E150422V1>.
- Baron, E. Jason, Ezra G. Goldstein, and Joseph P. Ryan.** 2021. "The Push for Racial Equity in Child Welfare: Can Blind Removals Reduce Disproportionality?" Unpublished.
- Baron, E. Jason, and Max Gross.** 2022. "Is there a Foster Care-to-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators." Unpublished.
- Barrat, Vanessa X., and BethAnn Berliner.** 2013. *The Invisible Achievement Gap, Part 1: Education Outcomes of Students in Foster Care in California's Public Schools.* San Francisco: WestEd.
- Barth, Richard P., Melissa Jonson-Reid, Johanna K.P. Greeson, Brett Drake, Jill Duerr Berrick, Antonio R. Garcia, Terry V. Shaw, and John R. Gyourko.** 2020. "Outcomes Following Child Welfare Services: What Are They and Do They Differ for Black Children?" *Journal of Public Child Welfare* 14 (5): 477–99.
- Berger, Lawrence M., Maria Cancian, Eunhee Han, Jennifer Noyes, and Vanessa Rios-Salas.** 2015. "Children's Academic Achievement and Foster Care." *Pediatrics* 135 (1): 109–16.
- Berger, Lawrence M., Sarah A. Font, Kristen S. Slack, and Jane Waldfogel.** 2017. "Income and Child Maltreatment in Unmarried Families: Evidence from the Earned Income Tax Credit." *Review of Economics of the Household* 15 (4): 1345–72.
- Berrick, Jill Duerr, and Julia Hernandez.** 2016. "Developing Consistent and Transparent Kinship Care Policy and Practice: State Mandated, Mediated, and Independent Care." *Children and Youth Services Review* 68: 24–33.
- Brehm, Margaret E.** 2021. "The Federal Adoption Tax Credit and Adoptions from Foster Care." *Journal of Human Resources* 56 (4): 1031–72.
- Bullinger, Lindsey R., Jason M. Lindo, and Jessamyn Schaller.** Forthcoming. "Economic Determinants of Child Maltreatment." In *Encyclopedia of Law and Economics*, edited by Alain Marciano and Giovanni Battista Ramello. 2nd ed. New York: Springer.
- Cancian, Maria, Mi-Youn Yang, and Kristen Shook Slack.** 2013. "The Effect of Additional Child Support Income on the Risk of Child Maltreatment." *Social Service Review* 87 (3): 417–37.
- Chadwick Center and Chapin Hall.** 2016. *Using Evidence to Accelerate the Safe and Effective Reduction of Congregate Care for Youth Involved with Child Welfare.* Chicago and San Diego: Chadwick Center and Chapin Hall.
- Chen, Ding-Geng, Mark F. Testa, David Ansong, and Kanisha C. Brevard.** 2020. "Evidence Building and Information Accumulation: Using the Bayesian Paradigm to Advance Child Welfare Intervention Research." *Journal of the Society for Social Work and Research* 11 (3): 483–98.
- Child Welfare Information Gateway.** 2018. *Home Study Requirements for Prospective Foster Parents.* Washington, DC: US Department of Health and Human Services, Children's Bureau.
- Child Welfare Information Gateway.** 2019a. *Background Checks for Prospective Foster, Adoptive, and Kinship Caregivers.* Washington, DC: US Department of Health and Human Services, Children's Bureau.
- Child Welfare Information Gateway.** 2019b. "What is Child Abuse and Neglect? Recognizing the Signs and Symptoms." Washington, DC: Children's Bureau, Administration for Children and Families.
- Children's Defense Fund.** 2020. *Implementing the Family First Prevention Services Act: A Technical Guide for Agencies, Policymakers, and Other Stakeholders.* Washington, DC: Children's Defense Fund.
- Courtney, Mark E., Nathanael J. Okpych, and Sunggeun (Ethan) Park.** 2018. *Report from CalYOUTH: Findings on the Relationship between Extended Foster Care and Youth's Outcomes at Age 21.* Chicago: Chapin Hall at the University of Chicago.
- Courtney, Mark E., Nathanael J. Okpych, Justin Harty, Huiling Feng, Sunggeun (Ethan) Park, Jenna Powers, Melanie Nadon, Dale J. Ditto, and Keunhye Park.** 2020. *Findings from the California Youth Transitions to Adulthood Study (CalYOUTH): Conditions of Youth at Age 23.* Chicago: Chapin Hall at the University of Chicago.

- Courtney, Mark E., Andrew Zinn, Heidi Johnson, and Karin E. Malm.** 2011. *Evaluation of the Massachusetts Adolescent Outreach Program for Youths in Intensive Foster Care: Final Report*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Detlaff, Alan J., and Reiko Boyd.** 2020. "Racial Disproportionality and Disparities in the Child Welfare System: Why Do They Exist, and What Can Be Done to Address Them?" *ANNALS of the American Academy of Political and Social Science* 692 (1): 253–74.
- Doyle Jr., Joseph J.** 2007a. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review* 97 (5): 1583–610.
- Doyle Jr., Joseph J.** 2007b. "Can't Buy Me Love? Subsidizing the Care of Related Children." *Journal of Public Economics* 91 (1–2): 281–304.
- Doyle Jr., Joseph J.** 2008. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." *Journal of Political Economy* 116 (4): 746–70.
- Doyle Jr., Joseph J.** 2013. "Causal Effects of Foster Care: An Instrumental-Variables Approach." *Children and Youth Services Review* 35 (7): 1143–51.
- Doyle Jr., Joseph J., and H. Elizabeth Peters.** 2007. "The Market for Foster Care: An Empirical Study of the Impact of Foster Care Subsidies." *Review of Economics of the Household* 5 (4): 329–51.
- Drake, Brett, Melissa Jonson-Reid, María Gandarilla Ocampo, Maria Morrison, and Darejan Dvalishvili.** 2020. "A Practical Framework for Considering the Use of Predictive Risk Modeling in Child Welfare." *ANNALS of the American Academy of Political and Social Science* 692 (1): 162–81.
- Dworsky, Amy, Laura Napolitano, and Mark E. Courtney.** 2013. "Homelessness During the Transition from Foster Care to Adulthood." *American Journal of Public Health* 103 (S2): 318–23.
- Fein, David J., and Wang S. Lee.** 2003. "The Impacts of Welfare Reform on Child Maltreatment in Delaware." *Children and Youth Services Review* 25 (1–2): 83–111.
- Finkelhor, David, Heather A. Turner, Anne Shattuck, and Sherry L. Hamby.** 2015. "Prevalence of Childhood Exposure to Violence, Crime, and Abuse: Results from the National Survey of Children's Exposure to Violence." *JAMA Pediatrics* 169 (8): 746–54.
- Fong, Kelley.** 2020. "Getting Eyes in the Home: Child Protective Services Investigations and State Surveillance of Family Life." *American Sociological Review* 85 (4): 610–38.
- Font, Sarah A., and Kathryn Maguire-Jack.** 2020. "The Scope, Nature and Causes of Child Abuse and Neglect." *ANNALS of the American Academy of Political and Social Science* 692 (1): 26–49.
- Garstka, Teri A., Crystal Collins-Camargo, Jennifer G. Hall, Melissa Neal, and Karl Ensign.** 2012. "Implementing Performance-Based Contracts and Quality Assurance Systems in Child Welfare Services: Results from a National Cross-Site Evaluation." *Journal of Public Child Welfare* 6 (1): 12–41.
- Goldhaber-Fiebert, Jeremy D., Kimberly S. Babiarz, Rachel L. Garfield, Fred Wulczyn, John Landsverk, Sarah M. Horwitz.** 2014. "Explaining Variations in State Foster Care Maintenance Rates and the Implications for Implementing New Evidence-Based Programs." *Children and Youth Services Review* 39: 183–206.
- Goldhaber-Fiebert, Jeremy D., and Lea Prince.** 2019. *Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County's Child Welfare Office*. Technical report. Accessed on 2022-02-26 from <https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-from-16-ACDHS-26-PredictiveRisk-Package-050119-FINAL-6.pdf>.
- Government Performance Lab.** 2019. *Transforming Service Delivery for Children, Youth, and Families in Rhode Island*. Providence: Harvard Kennedy School Government Performance Lab.
- Grimon, Marie-Pascale.** 2021. "Effects of the Child Protection System on Parents." Unpublished.
- Gross, Max, and E. Jason Baron.** Forthcoming. "Temporary Stays and Persistent Gains: The Causal Effects of Foster Care." *American Economic Journal: Applied Economics*.
- Hansen, Mary Eschelbach, and Bradley A. Hansen.** 2006. "The Economics of Adoption of Children from Foster Care." *Child Welfare* 85 (3): 559–83.
- Haskins, Ron.** 2020. "Child Welfare Financing: What Do We Fund, How, and What Could Be Improved?" *ANNALS of the American Academy of Political and Social Science* 692 (1): 50–67.
- Hayduk, Iryna.** 2017. "The Effect of Kinship Placement Laws on Foster Children's Well-Being." *B.E. Journal of Economic Analysis and Policy* 17 (1): 1–23.
- Hou, Claire.** 2021. "The Opioid Crisis and Foster Care Dynamics." Unpublished.
- Hubel, Grace S., Alayna Schreier, David J. Hansen, and Brian L. Wilcox.** 2013. "A Case Study of the Effects of Privatization of Child Welfare on Services for Children and Families: The Nebraska

- Experience." *Children and Youth Services Review* 35 (12): 2049–58.
- James Bell Associates.** 2019. *Special Topic Paper—Child Welfare Waiver Demonstrations and Evidence-Based Programs*. Washington, DC: Children's Bureau, Administration for Children and Families, U.S. Department of Health, and Human Services.
- Kim, Hyunil, Christopher Wildeman, Melissa Jonson-Reid, and Brett Drake.** 2017. "Lifetime Prevalence of Investigating Child Maltreatment Among US Children." *American Journal of Public Health* 107 (2): 274–80.
- Lawler, Patrick, and Jessica Foster.** 2013. "Making Performance-Based Contracting Work for Kids and Families." *Community Development Innovation Review* 9 (1).
- Lee, Bethany R., Charlotte L. Bright, Deborah V. Svoboda, Sunday Fakunmoju, and Richard P. Barth.** 2011. "Outcomes of Group Care for Youth: A Review of Comparative Studies." *Research on Social Work Practice* 21 (2): 177–89.
- Merritt, Darcey H.** 2020. "How Do Families Experience and Interact with CPS?" *ANNALS of the American Academy of Political and Social Science* 692 (1): 203–26.
- OECD.** 2013. OECD Family Database. <https://www.oecd.org/social/family/database.htm/> (accessed February 26, 2022).
- Okpych, Nathanael J., and Mark E. Courtney.** 2014. "Does Education Pay for Youth Formerly in Foster Care? Comparison of Employment Outcomes with a National Sample." *Children and Youth Services Review* 43: 18–28.
- Olds, David L., John Eckenrode, Charles R. Henderson Jr., Harriett Kitzman, Jane Powers, Robert Cole, Kimberly Sidora, Pamela Morris, Lisa M. Pettitt, and Dennis Luckey.** 1997. "Long-term Effects of Home Visitation on Maternal Life Course and Child Abuse and Neglect: Fifteen-Year Follow-Up of a Randomized Trial." *Journal of the American Medical Association* 278 (8): 637–43.
- Prettyman, Alexa.** 2020. "Happy 18th Birthday, Now Leave: The Hardships of Aging Out of Foster Care." Unpublished.
- Putnam-Hornstein, Emily, Barbara Needell, Bryn King, and Michelle Johnson-Motoyama.** 2013. "Racial and Ethnic Disparities: A Population-Based Examination of Risk Factors for Involvement with Child Protective Services." *Child Abuse and Neglect* 37 (1): 33–46.
- Rambachan, Ashesh, Jon Kleinberg, Sendhil Mullainathan, and Jens Ludwig.** 2020. "An Economic Approach to Regulating Algorithms." NBER Working Paper 27111.
- Reilly, Aaron, Heejin Kim, Lawrence M. Berger, Jessica Pac, and Kristen S. Slack.** 2021. "Performance-Based Contracts in Ongoing Work." Unpublished. Memorandum to WI Department of Children and Families.
- Roberts, Kelsey V.** 2019. "Foster Care and Child Welfare." PhD diss. Clemson University.
- Robinson-Cortes, Alejandro.** 2019. "Who Gets Placed Where and Why? An Empirical Framework for Foster Care Placement." Unpublished.
- Rouland, Bénédicte, and Rhema Vaithianathan.** 2018. "Cumulative Prevalence of Maltreatment Among New Zealand Children, 1998–2015." *American Journal of Public Health* 108 (4): 511–13.
- Rubin, David M., Evaline A. Alessandrini, Chris Feudtner, David S. Mandell, A. Russell Localio, and Trevor Hadley.** 2004. "Placement Stability and Mental Health Costs for Children in Foster Care." *Pediatrics* 113 (5): 1336–41.
- Sakai, Christina, Hua Lin, and Glenn Flores.** 2011. "Health Outcomes and Family Services in Kinship Care: Analysis of a National Sample of Children in the Child Welfare System." *Archives of Pediatrics and Adolescent Medicine* 165 (2): 159–65.
- Skemer, Melanie, and Erin Jacobs Valentine.** 2016. *Striving for Independence: Two-Year Impact Findings from the Youth Villages Transitional Living Evaluation*. New York: MDRC.
- Testa, Mark F., and David Kelly.** 2020. "The Evolution of Federal Child Welfare Policy through the Family First Prevention Services Act of 2018: Opportunities, Barriers, and Unintended Consequences." *ANNALS of the American Academy of Political and Social Science* 692 (1): 68–96.
- Turney, Kristin, and Christopher Wildeman.** 2016. "Mental and Physical Health of Children in Foster Care." *Pediatrics* 138 (5).
- US Bureau of Justice Statistics.** 2021. *Survey of Prison Inmates, United States, 2016*. Ann Arbor: Inter-university Consortium for Political and Social Research [distributor].
- US Department of Health and Human Services.** 2007. *Assessing Site Readiness: Considerations about Transitioning to a Privatized Child Welfare System*. Washington, DC: ASPE.
- US Department of Health and Human Services.** 2020a. *Adoption and Foster Care Analysis and Reporting System (AFCARS), Foster Care Files 1995-2019* [Dataset]. National Data Archive on Child Abuse and

- Neglect. Washington, DC: Administration for Children and Families, Administration on Children, Youth, and Families, Children's Bureau. <https://doi.org/10.34681/K17X-DD19/>.
- US Department of Health and Human Services.** 2020b. *Systemic Factors—Results from the CFSRs: 2015–2018*. Washington, DC: Administration for Children and Families, Administration on Children, Youth, and Families, Children's Bureau.
- US Department of Health and Human Services.** 2021. *Child Maltreatment 2019*. Washington, DC: Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau.
- US Government Accountability Office.** 2011. *Increased Data Sharing Could Improve Access to Benefits and Services*. Washington, DC: US Government Accountability Office.
- Vaithianathan, Rhema.** 2019. *Implementing a Child Welfare Decision Aide in Douglas County: Methodology Report*. Auckland: Centre for Social Data Analytics.
- Warburton, William P., Rebecca N. Warburton, Arthur Sweetman, and Clyde Hertzman.** 2014. "The Impact of Placing Adolescent Males into Foster Care on Education, Income Assistance, and Convictions." *Canadian Journal of Economics* 47 (1): 35–69.
- Westat, Chapin Hall Center for Children at the University of Chicago, and James Bell Associates.** 1998. *Evaluation of The New York City Home Builders Demonstration*. Washington, DC: ASPE.
- Wildeman, Christopher, and Natalia Emanuel.** 2014. "Cumulative Risks of Foster Care Placement by Age 18 for U.S. Children, 2000–2011." *PLOS ONE* 9 (3).
- Wulczyn, Fred.** 2020. "Foster Care in a Life Course Perspective." *ANNALS of the American Academy of Political and Social Science* 692 (1): 227–52.
- Wulczyn, Fred, Britany Orlebeke, Kristen Hislop, Florie Schmits, Jamie McClanahan, and Lilian Huang.** 2018. "The Dynamics of Foster Home Recruitment and Retention." Chicago: Chapin Hall at the University of Chicago.
- Yi, Youngmin, Frank R. Edwards, and Christopher Wildeman.** 2020. "Cumulative Prevalence of Confirmed Maltreatment and Foster Care Placement for US Children by Race/Ethnicity, 2011–2016." *American Journal of Public Health* 110 (5): 704–9.

Retrospectives

Joan Robinson on Karl Marx: “His Sense of Reality Is Far Stronger”

Carolina Alves

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please contact either Beatrice Cherrier, CNRS & CREST, ENSAE-Ecole Polytechnique (beatrice.cherrier@gmail.com) or Joseph Persky, University of Illinois at Chicago (jpersky@uic.edu).

Introduction

As the Great Depression unfolded in the early 1930s, the received economic orthodoxy was under challenge at Cambridge University, and pretty much everywhere else. The standard macroeconomic wisdom of the day was based on Say’s Law, that is, “production process (supply) generates the income necessary for the demand for these products” (Baumol 1999, p. 197), which offered no explanation for the collapse of output and commensurate rise in unemployment. The microeconomic wisdom of the day included the idea that each of the factors of production—in particular, labor and capital—was paid the marginal value of its output. At the time, the economic and political tectonic plates were shifting in Europe, which opened a space to question the orthodoxy of laissez-faire capitalism, and even to challenge what had often been the outright mockery of orthodox economists toward Marxism.

■ *Carolina Alves is the Joan Robinson Research Fellow in Heterodox Economics, University of Cambridge, Girton College, Cambridge, United Kingdom. Her email address is cca30@cam.ac.uk.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.36.2.247>.

At Cambridge, when John Maynard Keynes was formulating his challenge to the conventional wisdom in the early 1930s, he focused on how effective demand could fall short of what was needed to support full employment. But although Keynes was a fierce critic of the Say's Law aspect of orthodox theory, he in fact left most of the theoretical and methodological grounds of economics intact. Economist Piero Sraffa formed a discussion group known as the Cambridge Circus to discuss Keynes's work, but also to take a deep dive into the social problems associated with unemployment (Skidelsky 1992, p. 447). Once immersed in this rich and vibrant intellectual environment, Joan Robinson started a life-long journey to reconstruct economic theory. As a first step, she formulated *The Economics of Imperfect Competition* ([1933] 1969), providing a theoretical base for thinking about profits and wages in an economy in the absence of competition. Her Cambridge colleagues Sraffa and Michał Kalecki took a more radical stance, drawing upon a Marxist framework for analysis. They discussed crises and the impossibility of an automatic mechanism for re-establishing the equilibrium of the system, and placed the question of returns and distribution at the center of their analysis. They both took Marx's economics seriously, dissenting from the marginalist view that wages and returns to capital were determined by the marginal product of labor and capital, and instead beginning from the idea that capitalists would extract all they could from workers.¹

Robinson became both a friendly critic and supporter of Keynes, and in the early 1930s she provided elements for important developments in Keynes's thought (Harcourt and Kerr 2009, chap. 3; Robinson 1933a, 1933b). After Keynes published the *General Theory* in 1936, Robinson published her *Essays in the Theory of Employment* the following year, retrospectively considered the "very first post-Keynesian text" (Harcourt and King 1995, p. 32). By the end of the 1930s, Robinson wanted to bring academic and Marxian economics together, while also expecting Marx to "make her economics more 'real', to address the inequities of the capitalist world" (Harcourt and Kerr 2009, p. 34). Some of her like-minded contemporaries at Cambridge saw this as no more than a passing "flirtation" with Marx (Pasinetti 1987, p. 215). However, Cristina Marcuzzo (2003, p. 551), an expert in the Cambridge School of Economics, argues that Robinson had reached a fork in the road: that is, Robinson was unsatisfied with both orthodox economics and what would become known as Keynesian economics. On one hand, Robinson was still searching for a more realist theory of the rate of profit and income distribution. On the other hand, she had reservations about Keynes's concept of full employment and wondered about the nature of technical progress and a long-period theory within the Keynesian framework.

Of course, our current economic issues are not those of the Great Depression. However, we do live in a time where the United States, the United Kingdom, and other developed economies are experiencing a combination of high profit rates along with a falling labor share of income, and the high and rising inequality

¹See Kurz (2000) and Toporowski (2013, 2018) for an introduction to the works of Sraffa and Kalecki, respectively.

between capital and labor is staggering and undeniable. Decades after Robinson's time, critical issues of profit distribution and crisis mechanisms have not yet been fully addressed. Rather than simply attributing these patterns to shifts in the marginal product of labor and capital, one can instead investigate a heterodox view built on the bargaining power of capital and labor.

In this essay, I begin with some brief biographical notes on Robinson. I then focus on Robinson's 1942 *An Essay on Marxian Economics* (henceforth *Essay*). I revisit why Robinson felt the need to turn to Marx and which insights from Marxian economics she sought to incorporate into her later works. I comment on the legacy of the *Essay* and how it was received by some of her contemporaries. I conclude with some thoughts on how Robinson came to view herself as Marxist—although I would not label her work in that way.

Biographical Sketch

Joan Violet Robinson (1903–1983) was born in the leafy village of Camberley, Surrey, in England. Her father was Major General Sir Frederick Maurice and her mother, Helen Marsh, the daughter of a Professor and Master of Downing College, Cambridge. This rich family background led her to be privately educated, which gave her the foundation to apply and be admitted to the University of Cambridge in October 1922 via Girton College—Britain's first residential college for the degree-level education of women—despite the University not granting women degrees until 1948. Robinson concluded her studies in 1925, and by 1929 her academic life started to become intricately entwined with the story of Cambridge economics in the twentieth century, with first papers published in the early 1930s. However, Robinson's formal academic progress was murky. Despite giving supervision (small group teaching) at various Colleges in Cambridge, she never held a teaching Fellowship in any of those Colleges. She had to wait until 1934 for her first university post as Assistant Lecturer and then was finally made university Lecturer in 1937, Reader in 1949, and Professor in 1965.

Robinson's journey, even with such a rich family background, was no mean feat for a woman at the time. In Cambridge in the 1920s, women could not receive degrees, scholarships and fellowships were blocked to women, and only male graduate students had a seat at the administration table. The economics profession had no visible female role models and rarely were academic papers or books in economics attributed to women.² Robinson realized very early that her chances of being an academic were small. Reflecting on her position in the Cambridge milieu, she said she felt the “emotional conflicts of a hermaphrodite” (Robinson 1932c).

²When Robinson (1932b) published “Imperfect Competition and Falling Supply Price” in the *Economic Journal*, Gottfried Haberler, a Harvard Austrian-American economist, wrote to Richard Khan: “Who is Joan Robinson who wrote [in the last *Economic Journal*]? The Christian name sounds like a woman's but the article seems to me much too clever for a woman” (Haberler 1933).

However, she inherited a family trait of toughness and nonconformism seen in her father, who resigned his rank over principle after criticizing the UK government World War I policy (Gooch 1968), and her great-grandfather, a theologian who refused to believe in eternal damnation. She continued the family history of dissent as she pioneered unorthodox ideas in a bold, fearless, and outspoken manner.

Robinson shook the foundations of the economics of her time, introducing the negatively sloped marginal revenue curve, coming up with concepts such as monopsony (Thornton 2004), confronting the marginalist theory of distribution, and popularizing the unconventional economics ideas of John Maynard Keynes. In particular, *An Essay on Marxian Economics* (1942) is one of her three great works, the other two being *The Economics of Imperfect Competition* (1933) and *The Accumulation of Capital* (1956). It is also the least well-known, despite being the most important in terms of laying the foundations of her enduring challenge to the orthodox economics. If in 1933 Robinson offered an internal critique of Marshallian value theory, by 1942, with the *Essay*, she relied on Marxian insights to escape Marshallian orthodoxy. It is the story of how the originator of imperfect competition pushed further into a theory of exploitation.

Insights from Robinson's *Essay*

The orthodox economists have been much preoccupied with elegant elaboration of minor problems.... Marx's intellectual tools are far cruder, but his sense of reality is far stronger.

Robinson ([1942] 1966b)

Robinson was inspired by the conviction of Marx (and Keynes) that capitalism was run by capitalists and not consumers, which meant profit-making and accumulation were central goals for the capitalist class who “isn't at all interested in the full employment of workers” (Harcourt and King 1995, p. 34). Marx critically followed the step of classical political economists by looking into production and trade to understand the creation and distribution of wealth, with great attention to distributional conflicts. Some of his economic concepts may look as alien to a modern academic economist as they looked to Robinson in the late 1930s, but this did not dissuade Robinson (and should not dissuade us today). In his schema, capitalist and worker can be viewed as different agents with different interests, and their relationship is mediated by power. The monetary amount obtained from paying workers less than the amount they produce is called “surplus value,” and sits at the center of Marx's analysis of capitalism, underpinning this relationship of power. Here, we will explore the *Essay's* insights about profits, investment and unemployment, and exploitation using this framework.

The Rate of Profit

Robinson was dissatisfied with both Marshallian and Keynesian theories of profit, albeit for different reasons. The Marshallian view posited that the equilibrium

rate of profit would be a steady state with no net new investment, which seemed an unsatisfactory description of reality.³ Keynesian approaches of that time did allow for capital accumulation (as discussed in Robinson [1942] 1966b, p. xvii), but left other questions unaddressed. What if the accumulation of capitalists and consumption are limited by exploitation and the relationship determining it? What if the “mystery” of the constant relative shares of wages and profits in the product of industry within a country is due to trade unions forming a countervailing power to monopoly?

Marx enters this discussion through his scheme of how capitalism expands, which is the building block of his economic analysis of capitalism. Behind Marx’s scheme lies what he sees as the essential characteristic of the capitalist process—namely, capitalists producing and selling goods to obtain more money than they advance to buy means of production and hire workers. As Marx developed this scheme,⁴ it portrays the economy as two “departments” producing capital and consumption goods. The gross output of each department is given by

$$y = c + v + s$$

where c is constant capital (means of production used up in the production process), v is variable capital (wages advanced to workers), and s is surplus value. Marx refers to the value of labor-power as variable capital, because it is the part of capital used to buy the labor time that creates value. Only capitalists save, and they do so out of the appropriated surplus value. In each period, capitalists invest a certain proportion of the saved surplus value in the previous period, which provides more means of production in each department and more variable capital to hire workers. Total net production is distributed as wages to workers and surplus value to capitalists. Depreciation and wages are the only costs of production, while profit, interest, and rent are subdivisions of the surplus value. For simplicity, the discussion that follows will drop references to interest and rent, and so just use surplus value and profit interchangeably (unless explicitly stated otherwise).

This framework provides the basis for how Marx defines certain concepts. The division between payments to workers and to capitalists is determined by the rate of exploitation, which is defined here as the ratio of surplus value to variable capital: $e = s/v$.

³Robinson directly refers to the lack of a coherent theory of “normal” profits in the setting of the stationary state in Marshall. Robinson was challenged that she had mischaracterized Marshall (for example, Shove 1944), but in the preface of the second edition of the *Essay* in 1966, she maintained her initial claim, saying that the problem with Marshall is that he does not “give a coherent account of what determines the ‘normal’ rate of profit” (Robinson [1942] 1966b, p. xvi).

⁴For details of this discussion, see Marx’s analysis in the two volumes of *Capital* (1990, chap. 23; 1992 chap. 20–21). In Marx, there are two forms of social reproduction: simple and expanded. In the former, the whole surplus must be consumed by capitalists. In the latter, part of the surplus is invested, and it is where we have the issue of capital accumulation. For studies developing Marx’s expanded reproduction scheme, see Sardoni (2009) and Trigg (2006) as well as Robinson (1951).

In modern economics, the rate of profit can be viewed as a ratio: profit/capital stock. However, the rate of profit in Marx is the surplus relative to the sum of constant and variable capital: in the terms used here, $r = s/(c + v)$.⁵ Because Marxian profit is built on both surplus and variable capital (the wage bill), it will be affected by the level of exploitation.

In Marx's view, once the rate of exploitation is given, the rate of profit is derived from it, followed by the wage rate. If one asks what determines the wage rate, one must ask what determines the rate of exploitation first. In this way, the rate of exploitation takes us to the forces underlying class struggle in capitalism and thus outside the system of price and technology (Harris 1972).

For Robinson, Marx's concept of surplus value—and the related idea of exploitation—is the most striking difference between Marx and orthodox economics. At the time of the *Essay*, she accepts the idea of Marx's rate of exploitation and argues that the real wage determines the profit per worker employed, and the rate of profit on capital is then determined at the same time. This was clearly understandable in mid-twentieth-century heavy-industry Britain, where the idea that wages were paid by marginal product and profits tended to zero was hard to believe with any conviction. Yet Robinson would wrestle with this question for years to come, as she wondered whether there is a mechanism that establishes, given technical conditions, the rate of profit and then real wage rates emerge as residual; or vice versa, so that profit emerges as a residual (Robinson 1956, 1979b).

Marx also connects the rate of profit with what he calls the “organic composition of capital.” Under capitalism, there is a tendency for the productivity of labor to rise systematically. Marx tries to capture and measure this tendency through the concept of the organic composition of capital, defined in this framework as $k = c/v$. A rise in the organic composition of capital means that constant capital c is growing faster than variable capital v —or as we would describe it now, rising capital is raising profits but also displacing payments to workers, which leads to “a relatively redundant population of workers” known as the reserve army of labor (Marx 1990, p. 782).

Robinson seems to accept this idea of an organic composition of capital, but with reservations. She understands Marx to be arguing that changes in the conditions of production lead to a fall in profits: that is, an increase in k leads to a fall in r . This argument, in its simplest form, holds the rate of exploitation constant. Robinson clarifies that, with the rate of exploitation held constant, a rise in the organic composition of capital leads to a rise in productivity *and* levels of real wages. Furthermore, Robinson argues that this assumption of a fixed rate of exploitation is unfounded: instead, labor-saving techniques would reduce employment and raise output per head, “so that the rate of exploitation is raised and there is no reason to expect the rate of profit on capital to be falling—rather the reverse” (Robinson [1942] 1966b, p. xiii). In her view, Marx neither explained the idea that exploitation

⁵The Marxian rate of profit is more like a profit share, as s , c , and v are all flow variables. Robinson, however, finds it difficult to understand whether c and v are flows or stocks, which is one of her main critiques of Marxian economics.

could be treated as a constant nor noticed how a disproportionate rise in capital would lead to real wage increases.⁶

The *Essay* draws on Marx's insights for its critique of the orthodox theory of the rate of profit, and Robinson seemed fascinated by how these two different approaches lead to two very different ways of building a theory.⁷ Nonetheless, she did not accept that Marx had a coherent theory of profits until 1951, after reading Sraffa's introduction to *The Works and Correspondence of David Ricardo* (Harcourt and Kerr 2009, p. 47).

Investment and Unemployment

In the *Essay*, Robinson also expanded on the analysis of unemployment and imperfect competition, which she argued were modern developments that had "shattered the structure of orthodox doctrine and destroyed the complacency with which economists were wont to view the working of *laissez-faire* capitalism" (Robinson [1942] 1966b, p. xxii). In her view, Marx's scheme leads us to think about the possibility of an imbalance between supply and demand and the possibility of a crisis of disproportionality, overproduction (insufficient level of effective demand), and underconsumption (Marx's attack on Say's Law).⁸ Equally important, it takes our attention to investment demand and the inherent question of which mechanism can guarantee that capital accumulation takes place at the appropriate rate.

For Marx, investment represents additional demand for the output of the two departments of capital and consumption goods. For the flows between departments to balance, we must look into a unique proportional allocation of labor between the two departments, which implies that the "outlay" on labor needs to grow at the same rate in each department. Marx is aware of the difficulty of achieving such consistency for reasons I summarize in two points. First, there is no guarantee that the growth rate of employment will match that of the available labor force. Marx introduces the rate of employment as an additional variable linked to the reserve army of labor. For Robinson, the Keynesian theory of unemployment as the failure of effective demand can be complemented by Marx's idea of non-employment: the supply of labor growing faster than the number of jobs offered by the capitalist economy. Robinson also highlights that the reserve army is one of the mechanisms behind the determination of real wage, in the sense

⁶A number of Marx scholars would strongly disagree with Robinson that Marx did not explain his thinking here. Marx was aware that given that $r = s/c + v$, anything that reduces either c or v and anything that increases s is likely to increase r . Thus, he holds the rate of exploitation (defined as s/v) constant initially and shows the tendency of the rate of profit to fall, but then discusses possible counteracting factors to this tendency, such as super-exploitation of the work force (absolute surplus), cheapening of raw materials, wages, and goods through foreign trade, formation of joint stock companies and so on. More details about this can be found in Marx's discussion of value composition of capital (see also Fine and Saad-Filho 2010, pp. 95–96).

⁷See the online Appendix for a table with a systematic listing of differences between these approaches.

⁸In the 1942 *Essay*, Robinson thought that Marx's rejection of Say's Law was contradictory, as she interpreted that for Marx the law was inapplicable only during the crises. However, by the time of her 1953 essay "On Re-Reading Marx," she is convinced that for Marx, Say's Law never holds (Sardoni 1987, p. 66).

that bargaining strength between capitalist and workers is key in determining the movements of the level of wages. Thus, when labor-saving technical progress reduces the need for labor, it tends to lower the wage rate while raising the rate of exploitation.

Robinson seems enthralled by how Marx and Marxists discuss the interaction of unemployment and the rate of exploitation. She reminds us that in Marx, if the rate of exploitation rises because costs fall due to labor-saving technical progress while money-wages are constant, capitalists see a restriction upon the purchasing power of workers, so effective demand fails to expand with productive capacity. The labor share of output falls. Robinson seemed to embrace this notion of crises, shared of course by the Keynesians but based in a different logic.

The second concern raised by Marx about the balance between the investment and consumer departments begins with the insight that the private profit of the capitalist is the sole motivation for production. Similarly to Keynes, Marx divides all spending between consumption and investment (initially leaving aside government and foreign investment), but with a distinction between wage and profit income and the assumption that all wages are spent on consumption. If part of previous income is saved or if capitalists hoard money from their profits, total spending declines and there is overproduction, in which case the value of goods supplied exceeds the planned spending from the previous period. Overproduction, then, results in capitalists laying off workers and reducing investments.

Robinson's engagement with this point seems to show her heading to embrace the view that in capitalism money is put into production if some profit is expected. Otherwise, money is kept idle. Thus, there is a positive relationship between investment and expected profit, and accumulation is determined in Keynes and Marx by the energy of the capitalists (what Keynes called "animal spirits"). The pre-Keynesian economic orthodoxy, based on the assumption of full employment, was that the rate of saving governs investment, as mediated by an equilibrium real interest rate. Of course, if full employment is assumed, the theory becomes irrelevant not only for Robinson's ([1942] 1966b, p. 65) "modern world," but also for ours.

In any case, Robinson sees both Marx and Keynes distancing themselves from the orthodoxy due to their understanding that the determination of the rate of interest lies somewhere else than in the interaction between saving and investment: for Marx, in the bargaining power between lenders and borrowers; for Keynes, in the supply and demand for money and the issue of liquidity preference. For both of them, the interest rate can be an obstacle to capital accumulation. In Marx, the higher the interest rate, the larger is the share of surplus going to *rentiers*. In Keynes, low demand for money resulting from low activity and employment means that the rate of interest tends to fall (Robinson [1942] 1966b, pp. 68–69). Moreover, Keynes had shown that a rise in saving, which reduces consumption, leads also to reduction in income, which then stops increased saving from materializing. Thus, an initial rise in saving can lead to a fall in rate of investment, not an increase.

Although Robinson makes clear that Marx did not directly discuss effective demand, she sees in Marx many hints that could have led Marxists to elaborate a

theory of effective demand before Keynes did. She laments that the problem of effective demand does not arise in Marx, especially because he also “admits the divorce between decisions to save and decisions to invest, which, in Keynes’s system, appears as the root cause of crises and unemployment” (Robinson 1948, p. 141). For her, Marxists had much to learn from modern economics, at the time, when it came to effective demand and theory of employment, which could also provide a basis for the study of the laws of motion of capitalism (Robinson [1942] 1966b, p. xxiii).

Exploitation

Following in the footsteps of Marx, Robinson ([1942] 1966b, p. 3) sees capital accumulation and unprecedented wealth levels as the result of drawing out the productive power of combined and specialized labor, while workers producing the wealth do not share the benefits from the increase in their productive power. She emphasizes that when Marx is dividing the net product of industry into wages and surplus value (consisting of profit, interest, and rent), he is offering a theory of distribution that is also a theory of exploitation. Interestingly, however, Robinson rejected the idea that labor is the source of value, choosing instead to separate the theory of value from the theory of exploitation.⁹

Exploitation is an analytical category of significance for Robinson in both the 1933 *Imperfect Competition* and the 1942 *Essay*, albeit in related but different ways. In 1933, she showed that if imperfect competition takes place—which happens often—workers get paid less than the value of their marginal productivity. At the time she was an “analytical optimistic,” as she put it in her 1932 pamphlet *Economics Is a Serious Subject* (1932a), where she defends Marshall’s box of tools and the methodology of making unrealistic assumptions to give formal treatment to an economic problem (Marcuzzo 2010, p. 457). She defines exploitation as “payment to labor of less than its proper wage” (Robinson [1933] 1969, pp. 281–282), but class struggle and a discussion of the social relations at work in capitalism do not feature in her analysis.

In 1942, she reinstated the argument from 1933 that the main influence on the share of labor in the total output is the degree of imperfect competition in both the consumer good via monopoly and the labor market via monopsony. But her view is that the microeconomic focus on prices exposes us to minor defects in capitalism,

⁹Robinson’s interpretation and rejection of the labor value theory is also an outcome of dismissing Marx’s distinction between productive and circulation/exchange spheres. In Marx, there is a strict division between sphere of production (values) and circulation/exchange (prices). The discussion of the Marx’s reproduction scheme can be made in terms of either value or prices, and each of these concepts is equally relevant in its own sphere. Value allows us to assess the amount of the social product above the value of wages (surplus value/unpaid labor) that is appropriated by the owners of the means of production. Given the existence of surplus value for capitalists, prices perform the role of distributing this surplus among them in the sphere of exchange (Harris 1972). Robinson dismisses this division as a red herring. However, it is in the exchange under competitive conditions where capitalists try to expand their profits, which eventually leads to uniform this rate. That is why the rate of exploitation shows a relation between classes while the rate of profit shows us a relation between capitalists.

while Marx is concentrating on major issues by showing us the complexity of the social fabric and the different forces behind the distribution of the product between labor and capital. She brings issues related to the laws of motion of capitalism and crises to her analysis, coming from classical political economy. She seems to be heading toward a belief that distribution of income should be viewed as grounded in social and historical circumstances, not as a remuneration for scarce factors of production.¹⁰ For Robinson ([1942] 1966b, p. 80), both aspects are useful for a discussion of the nature of the system, but neither can tell us what really governs the margin of profit per unit of output.

By the time of *Essay*, Robinson had become even more careful about the tension between analytical and normative aspects of economics. She starts the *Essay* by making it clear that she wants to strip Marxian economics from ideology. She makes “amends” to Marxian concepts and ratios to make them not only understandable for an academic economist, but also empirically testable. However, her earlier “analytical optimistic” is tainted by a sober statement that orthodox economists assume the role of apologists of the system, while Marx wants to hasten its overthrowing. She argues that while Marx is aware of his purpose, orthodox economists are not, as “[t]hey wrote as they did because it seemed to them the only possible way to write, and they believed themselves to be endowed with scientific impartiality” (Robinson [1942] 1966b, p. 1).

Marx’s “awareness”—and bluntness about the utopia of scientific impartiality—gets Robinson’s attention when she highlights his point that not even algebraic formulae are innocent of political implications. Take, for example, Marx’s insistence in writing the rate of exploitation as surplus divided by variable capital (wage bill) s/v , not $s/(s+v)$. Robinson ([1942] 1966b, p. 22) writes:

The two formulations express precisely the same situation, but they imply two different attitudes to the capitalist process. The ratio s/v expresses the “real fact” of the “exclusion of labourer from the product” of his work, while the ratio $s/(s+v)$ presents the “false semblance of an association, in which the labourer and capitalist divide the product in proportion to the different elements which they respectively contribute towards its formation.”

In the *Essay*, Robinson is wondering where it could ever be possible to separate economic analysis from moral aspects and ideology. She wonders in a footnote about the “the transmutation of the notion of ‘exploitation’ which takes place under the influence of the modern theory.” From Marx’s views that labor is exploited because capital earns a net return, to the orthodox scheme where labor is exploited because it receives less than the wage given under perfect competition, she argues that in the

¹⁰This point was made earlier by Sraffa. In 1953, once fully clear about Sraffa’s point, Robinson would then criticize her *Imperfect Competition*, rejecting both capital as a factor of production and the rate of profit as the price which sets the quantity of capital employed (Marcuzzo 2003, p. 558).

orthodox scheme “moral and analytical considerations thus become inextricably confused” (Robinson [1942] 1966b, p. 77).

The *Essay*'s Legacy

[O]n the whole our regard for Marx is not increased but our regard for the old orthodoxy is somewhat diminished ...

Harrod (1942)

The *Essay*'s discussion is rooted in two fundamental differences Robinson sees between Marxian and orthodox economics. The first relates to whether one approaches capitalism as the natural order of things, as in the orthodox economics of the time, versus seeing it as a stage of development but neither the ideal nor the final stage, as in Marxian economics. The second relates to arguing in terms of a harmony of interests in society, as in the orthodox economics of the time, versus conceiving economic life in terms of a conflict of interests between property owners and workers with no property, as in Marxian economics. In orthodox economics, the shares of the social product among different classes are determined by inexorable natural laws. In Marxian economics, these shares are determined by the conflict between those who hope to gain and those who fear to lose.

But not only this noticeable contrast got Robinson's attention. For Robinson, Marxian economics provides a starting point for issues “where academic teaching was totally blank,” such as “growth and stagnation, technical progress and the demand for labor, the balance of sectors in an expanding economy” (Robinson 1962, p. 149). Even in the short run, she argues, Marx anticipated most of Keynes's theory, to the extent that starting with Marx would have saved Keynes “a lot of trouble” (Robinson 1964, p. 96).¹¹ For Keynes, the major feature of his theory of effective demand is that the economy can experience an underemployment equilibrium for long periods of time; for Marx, the “non-optimal outcomes” take the form of crises, especially overproduction crises. For Robinson, these two theories had close resemblance with what she called the modern theory of crisis. Marx's theory of the reserve army of unemployed labor and his theory of the relationship of capital-good to consumption-good industries (which shows the limitations of consumption) are what seem to have led her to expand on the Keynesian theories of unemployment and effective demand.

Marx helped Robinson to follow issues related to long-term accumulation that had only been touched on by Keynes (for example, Robinson 1980, p. 1475). As

¹¹ By 1964, Robinson considered Kalecki's theory of effective demand more satisfactory than Keynes's and attempted to enlarge the scope of Keynes's analysis to incorporate technical change, innovations, and changes in income distribution (Marcuzzo, 2010 p. 460). Later in life, Robinson would seek to make sure that Kalecki “received due recognition as the co-founder (with Keynes) of scientific macroeconomics” (Harcourt and King 1995, p. 32).

she wrestled with the problem of the inducement to invest, which for her was not wholly resolved, she asked which theory of investment was the most appropriate for a capitalist economy and wondered whether investment was primarily determined by long-run structural considerations. She was interested in how, for Marx, capital engages in the process of exploitation and how exploitation might make profit possible even over the long run. In short, an adequate concept of profit was necessary for extending Keynes's analysis of the long period. Robinson was starting a long path challenging the idea that the rate of profit is the price that sets the quantity of capital employed.

Marx's thinking on the total supply of capital and the rate of profit on capital was a key influence that led Robinson (1953) to tackle the problem of how capital is to be measured in the aggregate production function. Studying Marx helped Robinson to argue that in the long period we cannot avoid the question "what is the quantity of capital" (Marcuzzo 2018, pp. 124–125). By 1953, Robinson was pointing out "the neoclassical failure to distinguish between changes in the conditions of producing a given output when the quantity of capital is altered from changes in the value of that capital due to variations in wages and profits" (p. 125). Robinson's views in 1942 can be seen as the starting point of what became known as the "Cambridge capital theory controversies" (Groenewegen 2003), a lengthy dispute between economists at Cambridge University and a group based in Cambridge, Massachusetts, over whether capital could be meaningfully defined in an aggregate production function (for an overview in this journal, see Cohen and Harcourt 2003).

Marxian economics gave Robinson the insights and motivation she needed to move away from Marshallian economics and closer to a more classical approach (Robinson 1951, 1956, 1973c). Later on, in a move that would lead economists and historians of economics to consider her the midwife to post-Keynesian economics, she refused to follow Hicks's reconciliation of Keynes with classical theory under the IS-LM model, but rather "attempted to bridge Keynes, via Kalecki and Sraffa, to Marx and the classical authors" (Marcuzzo 2010, pp. 459–460).¹²

The reaction to Robinson's *Essay* was well mixed; it aroused controversy among orthodox and Marxian economists alike. Maurice Dobb, a Cambridge economist and historian who was part of Keynes's Cambridge milieu and a leading Marxian economist of his generation in Britain, had a continuous exchange with Robinson during the making of the *Essay*. He was skeptical about the idea of translating Marxian language to economics. He wrote to Robinson in 1941 saying "conventional ways of thoughts and modes of expression" are important parts of what one means, and "one is almost sure to miss, or even distort, the meaning in translation, as is so commonly done by historians of thought" (Dobb 1941). Dobb made clear to

¹²Robinson later labelled post-Keynesian economics as the true alternative to neoclassical economics. In her words, "the classical tradition, revived by Sraffa, which flows from Ricardo through Marx, diluted by Marshall and enriched by the analysis of effective demand of Keynes and Kalecki" (Robinson 1973b, p. xii). For an introduction to post-Keynesian economics, see King (2003) and Lavoie (2009).

Robinson that Marxian economic theory, or any theory for that matter, could not be reducible merely to deductive propositions.

Keynes responded to Robinson's effort with the remark: "there is something intrinsically boring in an attempt to make sense of what is in fact not sense" (Keynes 1942a). He mentioned how Marx would have faced fewer difficulties assuming that capital is not scarce so that "depreciation, if we forget about risk, would be the only cost of capital" (Keynes 1942b).

On the other side, Robinson's rejection of the theory of value caused her to be treated as an "enemy by the professed Marxists" (Robinson 1978, p. 276). Robinson could not see how applied research could be conducted in terms of Marx's system of labor values. In 1942, she argued that Marx's assertion that the value of a commodity consists of the labor time necessary to produce it is a purely dogmatic statement. In the 1966 edition of the *Essay*, she reinstates that the concept of value "in itself is quite devoid of operational meaning" and that it is a powerful metaphysical concept, while the rate of exploitation is not" (Robinson [1942] 1966b, p. xi). In contrast, she would state that the rate of exploitation is a matter of fact, not of definition (Robinson and Eatwell 1973, p. 29). Many Marxists at the time focused on her interpretation of the labor theory of value to argue that Robinson misread Marx and was not able to grasp the revolutionary character of Marxian economics and its critique of capitalism (Namboodiripad 1973; Jackson 1943). However, Robinson viewed the theory of value in the narrow sense of a theory of relative prices, which for her was not at "the heart of Marx's system" (Robinson 1950, p. 148). In her view, Marxians spent too much time on issues of prices, when a far more important issue was exploitation. Her interpretation was taken as an effort to decouple the theory of value from Marx's theory of class struggle, so as to translate Marx to mainstream economists.

Final Remarks

I have Marx in my bones.

Robinson (1953)

Robinson never fully accepted Marx's argument that exploitation was a cause of increasing misery of the industrial working class, and she did not follow the most revolutionary implications of Marxian economics. However, she found Marx's insights and ideas much more powerful for explaining reality than either the dominant Marshallian economics that she had been taught or the newly developing Keynesian economics of her time. She may not have adopted Marx's historical materialism ("dialectical") method where there is a continuous state of change resulting from interaction and conflicts, but Marx solidified in Robinson a path she saw starting with the Keynesian revolution regarding moving from the conception of equilibrium to the conception of history (Robinson 1973a).

Later in life, Robinson (1973b, p. x) would look back at her encounter with Marx in the 1930s and 1940s and say that for her, "the main message of Marx was

the need to think in terms of history, not of equilibrium.” Marx offers a long-run analysis with references to short-run declines in profit and investment, but he does not think in terms of equilibrium. Instead, Marx takes into account the dynamic character of the economy, not as a succession of static states, but rather as economic processes whose movements from period to period lead to ongoing changes and, thus, to an economic system that over time moves into disequilibrium. From Robinson’s perspective, this way of thinking is much more than noting that economic concepts and relationships are influenced by their historical and social setting. It is also about differentiating historical from logical time. As Robinson (1977, p. 57) later put it, “historical time moves from the dark past behind it into the unknown future in front,” while “logical time can be traced from left to right on the surface of a blackboard.” If we accept that the economy exists in time, our economic analysis must consider the difference between theories dealing with logical time and those dealing with historical time.

Although Robinson was still very much influenced by a tendency to moralize economics as a way of improving the world, which can be viewed as a legacy of the Marshallian economics she had been taught, she made very clear that Marx’s condemnation of the existing economic system was not merely about its moral repugnance. For Marx, periodic crisis and trade cycles were symptoms of a “deep-seated and progressive malady in the vitals of the system,” in contrast with the complacency of the orthodox academics who “preach the gloomy doctrine that all is for the best in the best of all *possible* worlds” (Robinson [1942] 1966b, p. 5). In this sense, for Robinson, Marx was ultimately encouraging, because he opens Pandora’s box and releases terror but also hope.

That said, Robinson’s radicalness seems to have roots before her encounter with Marx and even Keynes. Indeed, in her 1953 essay *On Re-reading Marx*, she wrote an *Open Letter from a Keynesian to a Marxist* stating that she was a “Left-wing Keynesian” long before the *General Theory* was published, and therefore also before her investigation of Marx. Her “pre-Marx” frustrations and critique of economics and capitalism are what seem to have made her state that she had Marx in her bones, without the need “of Hegelian stuff and nonsense” (Robinson 1953, p. 20).

I believe Robinson’s critique and praise of Marx were part of a genuine attempt to both bring Marx to economics and to make Marxian concepts intelligible and relevant for macroeconomists and applied macro research. For Robinson ([1942] 1966b, p. xx), “[i]t surely should be possible to adopt a language in which the two parties could talk to each other.” Some scholars like to highlight how Robinson tried to bring together macroeconomics and Marxian economic theory, stripping the ideological baggage of the latter. However, they often miss Robinson’s urge to disentangle ideology from *any* economic theory, including not just Marx but also Marshall or Keynes. She was always alert to distinctions between the intentions and the implications of an author. Robinson (1955, p. 71) wrote:

Marx’s analysis of capitalism shows its strong points, although his purpose was to attack it. Marshall’s argument inadvertently shows the wastefulness of

capitalism, although he meant to recommend it. Keynes in showing the need for remedies to the defects of capitalism also shows how dangerous the remedies may be.

In my view, Robinson examined Marx as one would approach any object of study, pointing out weaknesses and coming up with constructive criticisms while also acknowledging its strengths. She was interested in the progress of economic theory not in the abstract sense, but rather in terms of whether that theory was able to account for actual economic phenomena. I would not call Robinson a Marxist more than I would call her a Keynesian, or a member of any other school of thought. She was an original thinker who craved a better economics and was willing to look everywhere for the tools to build one.

■ *For helpful comments and discussion, I thank Cristina Marcuzzo, Angus Armstrong, Wendy Carlin, Jan Toporowski, Roberto Veneziani, Cahal Moran, Ingrid Kvangraven, Roberto Bigg, Alexander Douglas, and James Culham.*

References

- Baumol, William J.** 1999. "Retrospectives: Say's Law." *Journal of Economic Perspectives* 13 (1): 195–204.
- Dobb, Maurice.** 1941. Letter from Maurice Dobb to Joan Robinson. Joan Violet Robinson Papers, Cambridge University, jvr/vii/120/11–3.
- Cohen, Avi J., and G. C. Harcourt.** 2003. "Retrospectives: Whatever Happened to the Cambridge Capital Theory Controversies?" *Journal of Economic Perspectives* 17 (1): 199–214.
- Fine, Ben, and Saad-Filho, Alfredo.** 2010. *Marx's 'Capital'*. 5th ed. London: Pluto Press.
- Gooch, John.** 1968. "The Maurice Debate 1918." *Journal of Contemporary History* 3 (4): 211–28.
- Groenewegen, Peter.** 2003. "On Re-Reading Joan Robinson's 'On Re-Reading Marx.'" *Review of Political Economy* 15 (4): 509–19.
- Haberler, Gottfried.** 1933. Letter from Gottfried Haberler to Richard Kahn. Joan Violet Robinson Papers, Cambridge University, jvr/vii/171.
- Harcourt, G. C., and Prue Kerr.** 2009. *Joan Robinson*. London: Palgrave Macmillan.
- Harcourt, G. C., and John King.** 1995. "Talking About Joan Robinson: Geoff Harcourt in Conversation with John King." *Review of Social Economy* 53 (1): 31–64.
- Harris, Donald J.** 1972. "On Marx's Scheme of Reproduction and Accumulation." *Journal of Political Economy* 80 (3): 505–22.
- Harrod, R. F.** 1942. "'Economic Man': Review of Joan Robinson, *Essay on Marxian Economics*." *The Observer*, July 19.
- Jackson, T. A.** 1943. "Some Flowers for Marx's Grave." *The Plebs*, May.
- Keynes, John Maynard.** 1936. *The General Theory of Employment, Interest and Money*. Cambridge, UK: Cambridge University Press, 1973.
- Keynes, John Maynard.** 1942a. Letter from John M. Keynes to Joan Robinson. August 20, 1942. Modern Archives, King's College, Cambridge University, jmk/L/42/102–3.

- Keynes, John Maynard.** 1942b. Letter from John M. Keynes to Joan Robinson. August 21, 1942. Modern Archives, King's College, Cambridge University, jmk/CO/8/246.
- King, J. E., ed.** 2003. *The Elgar Companion to Post Keynesian Economics*. Cheltenham: Edward Elgar Publishing.
- Kurz, H. D., ed.** 2000. *Critical Essays on Piero Sraffa's Legacy in Economics*. Cambridge, UK: Cambridge University Press.
- Lavoie, Marc.** 2009. *An Introduction to Post-Keynesian Economics*. Basingstoke: Palgrave Macmillan.
- Marcuzzo, Maria Cristina.** 2003. "Joan Robinson and the Three Cambridge Revolutions." *Review of Political Economy* 15 (4): 545–60.
- Marcuzzo, Maria Cristina.** 2010. Review of *The Provocative Joan Robinson: The Making of a Cambridge Economist*. *Journal of Economic Literature* 48 (2): 456–60.
- Marcuzzo, Maria Cristina.** 2018. "Joan Robinson's Challenges on How to Construct a Post-Keynesian Economic Theory." *Annals of the Fondazione Luigi Einaudi* LII: 119–34.
- Marx, Karl.** 1990. *Capital: Critique of Political Economy*. Vol I. Translated by Ben Fowkes. New York: Penguin Classics.
- Marx, Karl.** 1992. *Capital: Critique of Political Economy*. Vol II. Translated by David Fernbach. London, New York: Penguin Classics.
- Namoodiripad, E. M. S.** 1973. "On Joan Robinson's Criticism of Marx." *Social Scientist* 1 (11): 18–31.
- Pasinetti, L. L.** 1987. "Joan Violet Robinson." In *The New Palgrave: A Dictionary of Economics*, Vol. 4, edited by J. Eatwell, M. Milgate, and P. Newman. London: Macmillan.
- Robinson, Joan.** 1932a. *Economics Is a Serious Subject: The Apologia of an Economist to the Mathematician, the Scientist and the Plain Man*. Cambridge: W. Heffer and Sons.
- Robinson, Joan.** 1932b. "Imperfect Competition and Falling Supply Price." *Economic Journal* 42 (168): 544–554.
- Robinson, Joan.** 1932c. Letter from Robinson to Austin. Austin Robinson Papers, Marshall Library of Economics, Cambridge University, 10/20–24/32, eagr/Box 8/2/1/13.
- Robinson, Joan.** 1933a. "A Parable on Savings and Investment." *Economica* 39: 75–84.
- Robinson, Joan.** 1933b. "The Theory of Money and the Analysis of Output." *Review of Economic Studies* 1 (1): 22–26.
- Robinson, Joan.** 1937. *Essays in the Theory of Employment*. London: Macmillan.
- Robinson, Joan.** 1948. "Marx and Keynes." In *Collected Economic Papers*, Vol. I. London: Basil Blackwell and Oxford, 133–145, 1966.
- Robinson, Joan.** 1950. "Karl Marx and the Close of His System." In *Collected Economic Papers*, Vol. I. London: Basil Blackwell & Oxford, 146–151, 1966.
- Robinson, Joan.** 1951. Introduction to *The Accumulation of Capital*, by Rosa Luxemburg. London: Routledge and Kegan Paul.
- Robinson, Joan.** 1953. *On Re-reading Marx*. Cambridge: Students' Bookshop.
- Robinson, Joan.** 1955. "Marx, Marshall and Keynes." In *Contributions to Modern Economics*. New York: Academic Press, 1978.
- Robinson, Joan.** 1956. *The Accumulation of Capital*. London: Macmillan.
- Robinson, Joan.** 1962. "Marxism: Religion and Science." In *Collected Economic Papers*, Vol. III. London: Basil Blackwell and Oxford, 92–9, 1965.
- Robinson, Joan.** 1964. "Kalecki and Keynes." In *Collected Economic Papers*, Vol. III. London: Basil Blackwell and Oxford, 1965.
- Robinson, Joan.** 1965. *Collected Economic Papers*, Vol. III. London: Basil Blackwell and Oxford.
- Robinson, Joan.** 1966a. *Collected Economic Papers*, Vol. I. London: Basil Blackwell and Oxford.
- Robinson, Joan.** 1966b. *An Essay on Marxian Economics*. 2nd ed. New York: Macmillan, 1942.
- Robinson, Joan.** 1969. *The Economics of Imperfect Competition*. 2nd ed. London: Palgrave Macmillan.
- Robinson, Joan, ed.** 1973a. *After Keynes*. Oxford: Basil Blackwell.
- Robinson, Joan.** 1973b. Foreword to *The Reconstruction of Political Economy: An Introduction to Post-Keynesian Economics*, by J. A. Kregel, ix–xiii. London: Palgrave Macmillan.
- Robinson, Joan.** 1973c. "Introduction to Reprint of 'On re-reading Marx.'" In *Collected Economic Papers*, Vol. IV. London: Basil Blackwell and Oxford, 247–48.
- Robinson, Joan.** 1975. *Collected Economic Papers*, Vol. II. London: Basil Blackwell and Oxford.
- Robinson, Joan.** 1977. "The Labour Theory of Value as an Analytical System." In *Collected Economic Papers*, Vol. V. London: Basil Blackwell & Oxford, 289–297, 1979.

- Robinson, Joan.** 1978b. "Formalism versus Dogma." *Journal of Contemporary Asia* 8 (3): 381–95.
- Robinson, Joan.** 1979a. *Collected Economic Papers*, Vol. V. London: Basil Blackwell and Oxford.
- Robinson, Joan.** 1979b. "Who Is a Marxist?" In *Collected Economic Papers*, Vol. V, 248–253. London: Basil Blackwell and Oxford.
- Robinson, Joan.** 1980. Review of *Marxian Political Economy: An Outline* by J. F. Becker. *American Journal of Sociology* 85 (6): 1476–8.
- Robinson, Joan, and John Eatwell.** 1973. *An Introduction to Modern Economics*. London: McGraw Hill.
- Sardoni, Claudio.** 1987. *Marx and Keynes on Economic Recessions: The Theory of Unemployment and Effective Demand*. Brighton: Wheatsheaf.
- Sardoni, Claudio.** 2009. "The Marxian Schemes of Reproduction and the Theory of Effective Demand." *Cambridge Journal of Economics* 33 (1) 161–73.
- Shove, G. F.** 1944. "Mrs. Robinson on Marxian Economics." *Economic Journal* 54 (213): 47–61.
- Skidelsky, Robert.** 1992. *John Maynard Keynes: The Economist as Saviour, 1920–1937*, Vol. II. New York: Penguin Press.
- Sraffa, Piero.** 1951. Introduction to *Works and Correspondence of David Ricardo*, Vol. I. Cambridge: Cambridge University Press.
- Thornton, Robert J.** 2004. "How Joan Robinson and B. L. Hallward Named Monopsony." *Journal of Economic Perspectives* 18 (2): 257–61.
- Toporowski, Jan.** 2013. *Michał Kalecki: An Intellectual Biography—Rendezvous in Cambridge, 1899-1939*. New York: Palgrave Macmillan.
- Toporowski, Jan.** 2018. *Michał Kalecki: An Intellectual Biography—By Intellect Alone, 1939-1970*. Cham, Switzerland: Palgrave Macmillan.
- Trigg, Andrew.** 2006. *Marxian Reproduction Schema: Money and Aggregate Demand in a Capitalist Economy*. London and New York: Routledge.

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

Laurence Boone, Joachim Fels, Òscar Jordà, Moritz Schularick, and Alan M. Taylor discuss *Debt: The Eye of the Storm* (Geneva Reports on the World Economy 24, Centre for Economic Policy Research, 2022, <https://voxeu.org/content/debt-eye-storm>). “The world economy today is characterised by a larger quantity of debt, relative to its level of income, than at any time in recorded history.... In the advanced economies, the average total debt of the household, business and government sectors has grown steadily and now exceeds 250% of GDP. In the emerging markets, the same average has risen more quickly in recent years and now stands at about 150% of GDP. The world is therefore highly levered in terms of its overall debt-to-income ratio ... The economic dislocations caused by the Covid-19 pandemic

■ *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.36.2.265>.

have jolted debt ratios even higher, raising even more concerns about future debt burdens and their potential dangers, but in many respects this crisis has only served to accelerate preexisting debt trends.... In sum, the picture that this report paints is one of cautious and perhaps unexpected optimism. As long as credit supply remains plentiful relative to debt issuance, and thus interest rates remain low, higher levels of debt are sustainable. We are not blind to the challenges policymakers will have to face. Nor are we blind to the possibility, in a world awash with debt, that negative shocks will generate more bouts of instability, which will inevitably spill over onto innocent bystanders in a globalised economy. However, the trends behind the oversupply of credit will likely continue for a long time. Debt should not be ignored. But neither should it be feared.”

The *World Development Report* is an annual flagship report of the World Bank, and the 2022 version focuses on “Finance for an Equitable Recovery” (February 2022, <https://www.worldbank.org/en/publication/wdr2022>). “Economic activity contracted in 2020 in about 90 percent of countries, exceeding the number of countries seeing such declines during two world wars, the Great Depression of the 1930s, the emerging economy debt crises of the 1980s, and the 2007–09 global financial crisis ... In 2020, the first year of the COVID-19 pandemic, the global economy shrank by approximately 3 percent, and global poverty increased for the first time in a generation.... As the economic effects of the pandemic continue, policy makers aim to strike a balance between providing enough support to mitigate the human costs of the crisis, while limiting the longer-term financial and macroeconomic risks that could emerge from higher debt levels resulting from the crisis. These risks are likely to arise more quickly in emerging economies and especially in low-income countries, where the public and private debt-carrying capacity is much lower than in advanced economies, and where economic conditions were, in many cases, challenging even before the pandemic.... Past crises have revealed that without a swift, comprehensive policy response, loan quality issues are likely to persist and worsen over time, as epitomized by the typical increase in loans to ‘zombie firms’—that is, loans to weak businesses that have little or no prospect of returning to health and fully paying off their debts. Continued extension and rolling over of loans to such firms (also known as evergreening) stifles economic growth by absorbing capital that would be better directed to loans for businesses with high productivity and growth potential.”

Ulrike Malmendier delivered the 2020 JEEA-FBBVA Lecture to the European Economic Association, now published as “Exposure, Experience, and Expertise: Why Personal Histories Matter in Economics” (*Journal of the European Economic Association*, December 2021, 19:6, pp. 2857–94). From the abstract: “Personal experiences of economic outcomes, from global financial crises to individual-level job losses, can shape individual beliefs, risk attitudes, and choices for years to come. A growing literature on *experience effects* shows that individuals act as if past outcomes that they experienced were overly likely to occur again, even if they are fully informed about the actual likelihood. This reaction to past experiences is long-lasting though it decays over time as individuals accumulate new experiences. Modern brain science helps understand these processes. Evidence on neural plasticity reveals that personal

experiences and learning alter the strength of neural connections and fine-tune the brain structure to those past experiences ('use-dependent brain')."

India's Ministry of Finance has published its annual *Economic Survey 2021-22* (<https://www.indiabudget.gov.in/economicsurvey>), with an overview of the current state of India's economy. "The default mode of policy-making in India and most of the world has traditionally been to rely on a pre-determined 'Waterfall' approach—an upfront analysis of the issue, detailed planning and finally meticulous implementation. This is the framework that underpins five-year plans and rigid urban master-plans. The problem is that the real world is a complex and unpredictable place buffeted by all kinds of random shocks and unintended consequences. The response of traditional economics was to create ever more detailed plans/regulations, and elaborate forecasting models despite more than adequate evidence that this did not improve outcomes. In his Nobel Prize acceptance speech, economist Friedrich Hayek dubbed this 'The Pretence of Knowledge'. This Economic Survey sets out to explain the alternative 'Agile' approach that informed India's economic response to the Covid-19 shock. This framework is based on feed-back loops, real-time monitoring of actual outcomes, flexible responses, safety-net buffers and so on. Planning matters in this framework but mostly for scenario analysis, identifying vulnerable sections, and understanding policy options rather than as a deterministic prediction of the flow of events.... Some form of feedback loop based policy-making was arguably always possible, but the Agile framework is particularly relevant today because of the explosion of real-time data that allows for constant monitoring. Such information includes GST [goods and services tax] collections, digital payments, satellite photographs, electricity production, cargo movements, internal/external trade, infrastructure roll-out, delivery of various schemes, mobility indicators, to name just a few."

Elroy Dimson, Paul Marsh, and Mike Staunton have co-authored the annual *Global Investment Returns Yearbook* published by the Credit Suisse Research Institute (February 2022, <https://www.credit-suisse.com/about-us/en/reports-research/csri.html>). This edition includes a special topics chapter on "Diversification." "Conventional wisdom is that a small number of stocks—say 10 to 20—is sufficient to provide market-mimicking returns. That interpretation is misleading ... Many more stocks are needed to create a well-diversified portfolio.... From 1980 onward, US investors made increasingly large investments in overseas equities. However, in risk-return terms, they would have been better off staying at home.... Moreover, this is before taking account of the higher costs of investing internationally in the earlier part of this period.... Stock-bond correlations have now been mostly negative in major world markets for some 20 years. This negative correlation means that stocks and bonds have served as a hedge for each other, enabling investors to increase stock allocations while still satisfying a portfolio risk budget.... In recent years, much research has focused on why the sign of the stock-bond correlation flipped in the late 1990s. What was different about the period before and afterwards? ... Despite the volume of research, neither theory nor empirical studies point to a single or clear explanation for the negative stock-bond correlation."

As a policy-focused complement to the “Symposium on Early Childhood Interventions” in this issue—a bipartisan group of academics under the moniker of the “AEI-Brookings Working Group on Childhood in the United States” has produced a report on “Rebalancing: Children First” (February 2022, <https://www.brookings.edu/research/rebalancing-children-first/>). “The working group proposes, in short, rewriting the generational contract. In 2019, the share of the federal budget spent on children was 9.2 percent and the share spent on the adult portions of Social Security, Medicare, and Medicaid was 45 percent.... This allocation is a statement of national priorities—priorities that the working group agrees need to change.”

Innovations in Money?

The Federal Reserve offers a framework, without committing to any policy choices, in a discussion on “Money and Payments: The U.S. Dollar in the Age of Digital Transformation” (January 2022, <https://www.federalreserve.gov/publications/files/money-and-payments-20220120.pdf>). “For the purpose of this paper, a CBDC [central bank digital currency] is defined as a digital liability of a central bank that is widely available to the general public. In this respect, it is analogous to a digital form of paper money.... A CBDC could potentially offer a range of benefits. For example, it could provide households and businesses a convenient, electronic form of central bank money, with the safety and liquidity that would entail; give entrepreneurs a platform on which to create new financial products and services; support faster and cheaper payments (including cross-border payments); and expand consumer access to the financial system. A CBDC could also pose certain risks and would raise a variety of important policy questions, including how it might affect financial-sector market structure, the cost and availability of credit, the safety and stability of the financial system, and the efficacy of monetary policy.... The Federal Reserve will continue to explore a wide range of design options for a CBDC. While no decisions have been made on whether to pursue a CBDC, analysis to date suggests that a potential U.S. CBDC, if one were created, would best serve the needs of the United States by being privacy-protected, intermediated, widely transferable, and identity-verified.”

Dirk Niepelt has edited an e-book of 19 short and readable essays on *Central Bank Digital Currency: Considerations, Projects, Outlook* (CEPR Press, November 2021, <https://voxeu.org/content/central-bank-digital-currency-considerations-projects-outlook>). Some of the essays focus on overall conceptual questions; others provide some detail on what CBDC experiments some central banks around the world are already carrying out. Niepelt writes in the introduction: “Many central banks have concluded that private sector initiatives have taken the status quo option off the table. And as a consequence, more and more monetary authorities have morphed from observers into active contributors, albeit often sceptical ones. While the discussion about the ‘right’ CBDC choices—no, or yes and how—is far from settled, the

arguments have become sharper and the trade-offs clearer. It has become evident that the implications of CBDC extend far beyond the realms of payments, monetary policy and financial stability. And consequently, there is growing awareness that parliaments and voters—not only central banks—should actively join the debate.”

The President’s Working Group on Financial Markets (PWG), joined by the Federal Deposit Insurance Corporation (FDIC) and the Office of the Comptroller of the Currency (OCC), released a “Report on Stablecoins” (November 2021, https://home.treasury.gov/system/files/136/StableCoinReport_Nov1_508.pdf). “Stablecoins are digital assets that are designed to maintain a stable value relative to a national currency or other reference assets. Today, stablecoins are primarily used in the United States to facilitate trading, lending, or borrowing of other digital assets, predominantly on or through digital asset trading platforms. Proponents believe stablecoins could become widely used by households and businesses as a means of payment. If well-designed and appropriately regulated, stablecoins could support faster, more efficient, and more inclusive payments options.... Speculative digital asset trading, which may involve the use of stablecoins to move easily between digital asset platforms or in decentralized finance (DeFi) arrangements, presents risks related to market integrity and investor protection. These market integrity and investor protection risks encompass possible fraud and misconduct in digital asset trading, including market manipulation, insider trading, and front running, as well as a lack of trading or price transparency. Where these activities involve complex relationships or significant amounts of leverage, there may also be risks to the broader financial system.... Stablecoins also pose illicit finance concerns and risks to financial integrity, including concerns related to compliance with rules governing anti-money laundering (AML) and countering the financing of terrorism (CFT) and proliferation.”

Interviews

Noah Smith serves as interlocutor in “Interview: Emi Nakamura, macroeconomist” (Noahpinion, February 21, 2022, <https://noahpinion.substack.com/p/interview-emi-nakamura-macroeconomist>). “The recent increase in inflation is much more than historical experience would have predicted (which is about an increase in inflation of 1/3% for every 1% decrease in unemployment). I think several factors have been important. First, after a long hiatus from playing a major role in inflation, supply shocks are back! The most dramatic of these is the disruptions to the labor market. US labor force participation is down by roughly 1.5%, and so far the decline is pretty persistent.... There have also been other important supply shocks: it’s more expensive to operate a daycare or a factory than it used to be due to safety restrictions due to COVID. It used to be hard to come up with good examples of negative supply shocks in teaching undergraduate economics classes, but COVID certainly counts as one! Second, there has been a historic shift in demand from services to goods. In the Great Recession, the fraction of expenditures spent

on goods fell. The opposite happened during COVID: the fraction of spending on goods rose pretty dramatically. This is another tectonic shift in the economy that I think is putting enormous pressure on supply chains among other things.... Third, there has been a very rapid recovery and a lot of government support for spending. Households have a huge buildup in savings, and spending this down is no doubt contributing to demand.... One thing that hasn't contributed much to inflation so far is an unhinging of longer run inflation expectations. Both survey and market-based measures of longer run inflation expectations look pretty stable ... It's one of the Fed's primary goals these days to keep it that way."

David A. Price conducts an "Interview" with Pinelopi Goldberg, subtitled "On developing countries, measuring economies by satellite, and the learning crisis" (*Econ Focus*, Federal Reserve Bank of Richmond, Winter 2022, pp. 22–26, https://www.richmondfed.org/publications/research/econ_focus/2022/q1_interview). At one point, Goldberg discusses her research (with Tristan Reed) on why COVID death rates seem so much lower in low-income countries: "Most people's reaction was that this result was just because poor countries are not connected, so COVID-19 had not arrived there yet.... [But] capitals of low-income countries are not as isolated as people think; many of these cities are global cities. They are connected to the rest of the world.... Another reaction was that this was all measurement error.... But the differences in deaths are huge—orders of magnitude apart. Just to give you one striking example, in the United States right now, the deaths per million are around 2,500. In Nigeria, the number is 14; in India, it's 340. And it's not easy to hide deaths. Yes, there is measurement error—probably deaths and hospitalizations are much higher in low-income countries than the statistics show—but still, there is a big difference between low-income countries and richer ones. I think there are three reasons at work.... First, everyone agrees that two of the risk factors for a serious reaction to COVID-19 leading to hospitalization and death are age and obesity. The age distribution in many low-income countries is very different from that in the United States. To mention a striking case, in Niger, the median age is 15; there, COVID-19 would probably not have very severe health effects on the population. On top of that, in low-income settings, obesity is much lower.... In addition, many epidemiologists talk about what they call "trained immunity" for low-income countries. The idea is that people in those countries are exposed to disease all the time, so their immune systems have learned how to cope. An alternative interpretation is that there has been selection; the ones who have managed to survive the various diseases they've been exposed to have very strong immune systems.... It's still the case that the poorer the country, the lower the per capita COVID-19 deaths so far. We'll see whether this holds in the future."

Ilan Goldfajn and Eduardo Levy Yeyati have edited an e-book, *Latin America: The Post-Pandemic Decade—Conversations with 16 Latin American Economists* (CEPR Press, December 2021, <https://voxeu.org/content/latin-america-post-pandemic-decade-conversations-14-latin-american-economists>). As one example, here are comments from Ricardo Hausmann: "Over the past 60 years, the [Latin America] region has not shown a capacity to narrow the huge income gap it has with the advanced

countries ... The region's income per capita at market prices is less than 1/7th that of the United States, 1/4th when adjusted for purchasing power parity.... This lack of progress in closing the income gap is surprising in light of the fact that gaps in education, health, life expectancy, infant mortality, urbanisation, fertility rates and female labour force participation have narrowed dramatically or even reversed, while gaps in investment effort are either small or negative.... Policies have also become more market friendly: inflation is way down in the single digits (except in Venezuela and Argentina), credit ratings have improved, trade has been liberalised, public enterprises have been privatised and many other indicators of market-friendly structural policies have all been moving in the right direction ... My interpretation of this state of affairs is that there is a growing technology gap: Latin America is particularly bad at adopting and adapting technology.... For the region as a whole, the rate of patenting is 1/70th that of the United States, with the best performers—Brazil and Chile—at about 1/40th of US levels. This is not a typical feature of middle-income countries: China's patenting rate per capita is higher than the United States and Korea holds the world patenting record per capita, while Turkey, Eastern Europe and the former Soviet Union dwarf the best Latin American performers.... A final piece of evidence is the dearth of new exports in Latin America: while the export basket of fast-growing countries in East Asia and Eastern Europe shows rapid diversification and sophistication—from garments, to electronics, to cars, to machinery, to chemicals and beyond—Latin America has been stuck in a narrow set of exports. Even Latin America's own positive deviance—blueberries, soybeans, avocados and other fruits—speak about technological developments adapted to local conditions that allowed the region to deploy physical and human capital into new 'ideas'."

Discussion Starters

Keith Fuglie, Jeremy Jelliffe, and Stephen Morgan of the Economic Research Service at the US Department of Agriculture discuss "Slowing Productivity Reduces Growth in Global Agricultural Output" (*Amber Waves*, December 28, 2021, <https://www.ers.usda.gov/amber-waves/2021/december/slowing-productivity-reduces-growth-in-global-agricultural-output/>). "After the 'Green Revolution' of the 1960s and 1970s increased crop yields, growth rates in agricultural output peaked and then started declining until the 1990s. In the 1990s and 2000s, growth rates in global agricultural output began trending upward again. Driven by more rapid TFP [total factor productivity] growth, world agricultural output increased fast enough to meet food demand from growing global populations and incomes, and reduced the need to use more land, labor, and other resources. However, ... global agricultural productivity growth slowed in the 2010s. Output grew at an average annual rate of 2.08 percent during 2011–19 compared with 2.68 percent during 2001–10 ... Average TFP growth during these decades fell from 1.96 percent annually to 1.31 percent annually.... At the same time, the rate at which agricultural land was

expanding increased.... The slowdown in agricultural output growth over the past decade occurred primarily in developing countries ... The growth rate in agricultural production in developing countries declined from an average of 3.56 percent in the 2000s to 2.37 percent in the 2010s. A decrease in the rate of TFP growth was responsible for most of this decline.”

Dorte Verner, Nanna Roos, Afton Halloran, Glenn Surabian, Edinaldo Tebaldi, Maximillian Ashwill, Saleema Vellani, and Yasuo Konishi have written a book on *Insect and Hydroponic Farming in Africa: The New Circular Food Economy* (World Bank, December 2021, <https://openknowledge.worldbank.org/handle/10986/36401>). From the Executive Summary: “While current agri-food production models rely on abundant supplies of water, energy, and arable land and generate significant greenhouse gas emissions in addition to forest and biodiversity loss, past practices point toward more affordable and sustainable paths.... Frontier agriculture includes insect farming and hydroponic farming ... Insect farming is the process of producing insects for human food and animal feed, and hydroponic farming is the process of growing crops in nutrient-rich water solutions instead of soil. These technologies do not require great access to land, water, or wealth—all limiting factors in Africa ... The technologies use organic waste, including agricultural or certain industrial waste, to quickly produce nutritious and protein-rich foods for humans, fish, and livestock and biofertilizers for soils. This improves food and nutrition security while reducing waste, strengthening national accounts, replenishing the environment, saving hard currency reserves by decreasing food and feed imports, and promoting green, resilient, and inclusive development.... Within a year, African insect farming can generate crude protein worth up to US\$2.6 billion and biofertilizers worth up to US\$19.4 billion. That is enough protein meal to meet up to 14 percent of the crude protein needed to rear all the pigs, goats, fish, and chickens in Africa, according to the report’s modeling of the annual production of black soldier fly larvae (BSFL) in Africa.”