SCHOLARLY PERSPECTIVES

UCLA SCHOOL OF LAW
TABLE OF CONTENTS

ITERATIVE FEDERALISM AND CLIMATE CHANGE 1
Ann E. Carlson

GOOGLE BOOK SEARCH IS NOT FAIR USE 15
Douglas Lichtman

A LABOR LAW FOR THE DIGITAL ERA 39
Katherine V.W. Stone

WAR AND TAXES 66
Steven A. Bank and Kirk J. Stark
Ann Carlson is professor of law and the inaugural faculty director of the Emmett Center on Climate Change and the Environment at UCLA School of Law. She is also on the faculty of the UCLA Institute of the Environment. Professor Carlson’s scholarship in environmental law focuses on climate change law and policy, federalism and the role social norms play in affecting environmentally cooperative behavior. Her recent work involves analyzing unusual models of environmental federalism, with a focus on the unique role California plays in regulating mobile source emissions, including greenhouse gas emissions, under the Clean Air Act. She has also written on the legal and political obstacles utilities will face in cutting greenhouse gas emissions and on the threat of heat waves and climate change. She is a frequent commentator and speaker on environmental issues, particularly on climate change. Professor Carlson’s article, “Takings on the Ground” was selected in 2003 by the Land Use and Environmental Law Review as one of the top ten environmental articles of the year. She is co-author (with Daniel Farber and Jody Freeman) of ENVIRONMENTAL LAW (7th Ed.). Professor Carlson teaches Property, Environmental Law and Climate Change Law and Policy and was the recipient of the 2006 Rutter Award for Excellence in Teaching. She served as the law school’s academic associate dean from 2004–2006. Carlson received her J.D. magna cum laude from Harvard Law School in 1989 and her B.A., magna cum laude, from the University of California at Santa Barbara in 1982.
ITERATIVE FEDERALISM AND CLIMATE CHANGE

Ann E. Carlson

While the federal government has remained on the sidelines, a number of states have produced interesting and innovative programs to reduce greenhouse gas emissions. This Article suggests that though states deserve credit for their climate change leadership, at least two of the most significant state initiatives—California’s greenhouse gas mobile source emissions standards and the Regional Greenhouse Gas Initiative—would not have occurred but for the backdrop of federal law. Indeed these two state initiatives are the end product of what the Article terms “iterative federalism.” Under iterative federalism schemes, federal law singles out a state or particular group of states for special regulatory power rather than treating all fifty states as legally homogeneous. The result has been an iterative pattern of regulatory innovation, under which the special state “super-regulator” moves to regulate, followed by the federal government, followed in turn by more regulatory innovation by the state super-regulator and so forth.

These schemes of iterative federalism not only produce regulatory innovation and significant environmental success but also shed new light on the long-standing debate about which level of government, state or federal, is the appropriate locus of regulatory power in environmental policy making. Iterative federalism retains some of the chief benefits of devolution—policy experimentation, respect for local preferences and the avoidance of potentially expensive and untested federal mandates. Yet iterative federalism schemes quite effectively address interstate externalities, national product market economies of scale and the race to the bottom often feared by advocates of national environmental policy making. And through the creative deployment of federal law, iterative federalism has bolstered innovative state environmental leadership.

The federal government has remained on the sidelines for the past eight years as scientific evidence has mounted that the earth is warming at an alarming pace. Scientists believe with near certainty that human activity is a central cause of that warming, primarily through the burning of fossil fuels.¹

Though the federal government has failed to act to regulate greenhouse gas emissions, over the past several years the United States has hardly been idle. Instead,
a surprisingly large number of states have stepped in to fill the policy void.² States have enacted renewable portfolio standards; created incentives for carbon capture and sequestration; mandated energy efficiency standards; and established public benefit funds to support energy efficiency and renewable energy.³ Some states have gone even further, enacting overall greenhouse gas emissions caps,⁴ adopting greenhouse gas emissions standards for new automobiles,⁵ and capping utility emissions.⁶

But the standard account of state action on climate change misses a large part of the story. Conventional thinking emphasizes how the states have partly filled the regulatory voids created by federal inaction. This thinking, however, misses the critical ways in which the most innovative state responses to climate change are neither simply the product of state regulation nor exclusively federal. Instead, they are the results of repeated, sustained and dynamic lawmaking efforts that involve both levels of government.

“Iterative federalism,” I argue, is in fact the best label for describing two of the most significant climate change initiatives to come from the states—California’s mobile source emissions standards and the Regional Greenhouse Gas Initiative (RGGI). While the national government has failed to lead, the federal government's long history of environmental policy making has shaped and enabled state responses to climate change. But my claim goes further than this. I argue that without the role played by the federal government in enabling the particular states or regions to act, these two state climate change initiatives would literally not have occurred. To understand how and why, one must look not just at the inactive federal government or its activist state counterparts but at the interaction between state and federal law, at iterative federalism.

First, a clarification. In identifying and analyzing examples of iterative federalism, I mean to distinguish iterative federalism from federalism schemes that involve areas where state and federal areas of jurisdiction merely overlap through independent exercises of policy making authority.⁷ Instead, my focus is on schemes of federalism where federal law quite consciously designates a particular and distinct state or group of states to regulate and uses that regulatory arrangement to enhance compliance with federal standards.

The examples I identify of iterative environmental federalism share two characteristics. To start, rather than treating all fifty states as legally homogenous, federal law has singled out a state or group of states for special regulatory power. California’s special
status in regulating automobile emissions under the Clean Air Act (CAA)—which it used to enact its greenhouse gas emissions legislation—provides one example. The establishment of the Ozone Transport Commission (OTC) with its ten Northeastern state membership in the 1990 amendments to the Clean Air Act—out of which grew RGGI—provides another. Second, federal law undergirds this special state regulatory power by requiring the state regulator to comply with national environmental standards. Out of this dynamic, in which the federal government has not acted itself but has quasi-deputized a state or region to act while simultaneously regulating its actions, a quite interesting version of federalism emerges. Under it one level of government—either the singled-out state actor or the national government—moves to regulate in a particular environmental policy area. The initial policy making then triggers a series of iterations adopted in turn by the higher/lower level of government and then back to the policy originator and so forth.

In both the California and OTC examples, the regulatory exceptionalism contained in the Clean Air Act has produced a robust series of policy iterations that has resulted not only in large air pollution reductions but has also expanded the initial regulatory experimentation beyond the borders of the super-regulator jurisdictions. And both iterative federalism schemes have produced two ambitious and interesting legislative initiatives to reduce carbon emissions. California has enacted greenhouse gas emissions standards for passenger automobiles and the OTC states have entered a memorandum of understanding to impose a cap and trade scheme on electric utilities to regulate carbon dioxide emissions. And just as the air pollution iterations have expanded beyond the super-regulator’s borders, it is likely that the climate change regulatory schemes will do so as well.

In order to put the regulatory efforts of California and the Ozone Transport Commission into context, a bit of brief background about the operation of the Clean Air Act is necessary. The basic framework for controlling air pollution since the enactment of the modern Clean Air Act in 1970 is one of cooperative federalism: the Environmental Protection Agency, through its delegated authority under the Act, has issued National Ambient Air Quality Standards (NAAQS) for harmful air pollutants. The EPA has designated six “criteria” pollutants for which NAAQS are established, including carbon monoxide, lead, nitrogen dioxide, ozone and particulate matter. The standards (set as allowable parts per million) are designed to protect human health and, in some instances, the physical environment.

The CAA delegates to states the authority to implement the NAAQS through the
adoption of State Implementation Plans (SIPs). States are given a fair amount of discretion to devise their plans in a manner that takes into account local geographical and economic conditions, voter preferences and the like, so long as a state's SIP contains measures that will either attain or maintain the NAAQS and, importantly, mitigate the transport of interstate air pollution. Though states were supposed to meet the NAAQS by 1975, Congress has twice extended the NAAQS deadlines and numerous areas of the country—principally the cities of the Northeast, parts of Texas and California—remain out of compliance for ozone and particulate matter. In addition to the central features of the CAA, two provisions are of special interest to my claims here. One grants California special authority to regulate motor vehicle standards. The other provision establishes the Northeast's Ozone Transport Commission. I describe these special provisions and the resulting regulatory activity next.

California is the only state in the country authorized to enact its own vehicle emissions standards. All other states are preempted from doing so under the federal Clean Air Act. Other states can, however, opt into the California standards or remain subject to federal standards, which are typically less stringent than California’s.

The California experience as a “super-regulator” under a scheme of iterative federalism has been a rather remarkable one, leading to at least nine separate iterations of emissions standards. Typically, the pattern has been that California enacts ambitious motor vehicle standards and within a year or two the federal government follows suit. A number of states, typically in the Northeast and Pacific Northwest, have opted into the California standards.

The various iterations include the first tail pipe standards in the mid-1960s, which were tightened numerous times between that time and 1990. Over that twenty-five year period California’s efforts led to standards that cut nitrous oxide, carbon monoxide and hydrocarbons emissions by more than 90 percent. Post 1990 California shifted its mode of regulation to create extremely low emissions vehicles based on fleet standards. The regulatory program has been so successful that the state’s Air Resources Board chairman describes them as follows: “We’ve seen the near impossible accomplished with gasoline vehicles: zero evaporative emissions, exceedingly clean exhaust—cleaner, in some cases, than the outside air entering the cabin for ventilation purposes and emission control systems that are twice as durable as their conventional forebearers, forecasted to last an astonishing 150,000 miles.”
Slightly less stringent low emissions vehicle standards—modeled after the California program—have been adopted at the federal level.

While California has been the first mover on mobile source emissions standards, the northeastern part of the country has quite successfully experimented with regulating air pollution by adopting cap and trade schemes. Generally speaking these schemes set an overall cap on a particular pollutant and then allocate to major polluters allowances or credits. Each credit, typically, allows its holder to emit one ton of the regulated pollutant. If a polluter pollutes less than the amount its credits allow, the polluter can sell excess credits to polluters who need more. If a polluter lacks sufficient numbers of credits it can purchase unused credits.

Unlike with mobile source emissions, the first level of government to enact a cap and trade program was the federal government in passing the 1990 Acid Rain Program. The Acid Rain Program regulates sulfur dioxide. Based on that experience and under authority granted to them by a separate provision of the Clean Air Act, 11 Northeastern states and the District of Columbia enacted a cap and trade program to regulate ozone pollution. These states, acting under the auspices of the Ozone Transport Commission, worked together in an attempt to combat cross border ozone pollution. The cap and trade scheme they adopted was a smashing success by virtually all measures. Each year of the program—from 1999 through 2002—saw double digit declines in the percent of unused allowances below the total cap (20 percent in 1999, 11 percent in 2000, 12 percent in 2001, and 11 percent in 2002). Moreover, emissions fell during peak ozone season and on particularly hot days (a problem for smog formation not only because of the temperature but because electricity generation soars as temperatures increase). The emissions trading program also achieved almost perfect compliance rates and very little “leakage”—emissions migrating from a regulated area to a non-regulated area—as a result of the program. The program was so successful that it led to a third iteration, called the NOx Budget Trading Program. The NOx Budget Trading Program, adopted by the EPA, used the Ozone Transport Commission’s cap and trade program and expanded it to include eleven states in addition to the Northeastern participants, many of them Midwestern and Southern states that have caused significant cross border pollution in the Northeast. Preliminary results show that the new program has also succeeded in reducing ozone pollution by large amounts.
The deployment of federal law to create “super-regulators” has succeeded in creating a particularly robust and dynamic series of iterations that have resulted in two significant achievements. First, the California and OTC provisions have led to large reductions in air pollution. Second, the provisions have created regulatory capacity in California and the OTC states that have led to major state initiatives on climate change, more thoroughgoing and significant than the states would have been likely to produce without the federal role.

California has used its special authority to enact the country’s first greenhouse gas emissions standards for passenger automobiles. These standards are modeled directly on the state’s most recent air pollution regulations establishing extremely low emissions vehicle tiers. And the state’s influence has expanded well beyond its borders: at least fifteen states have indicated that if the California standards are allowed to go into effect they will enact them. The Northeastern states have used their regulatory expertise to enact the first greenhouse gas emissions cap and trade scheme in the country. The greenhouse gas emissions scheme looks almost identical in operation to the cap and trade scheme the same states adopted to tackle ozone pollution. Other states are using the Northeastern state experience to craft their own cap and trade programs, including California.

An examination of iterative federalism schemes contributes to ongoing theoretical debates about federalism within the environmental context. Two key claims have emerged in legal scholarship about environmental federalism. The first claim is that a flurry of state environmental regulatory activity can lead to uniform federal legislation as a result of pressure from the regulated community. The second argues that states are more likely to produce efficient levels of environmental regulation because of interstate competition for capital and residents. Here I identify a third pattern.

In a significant and widely cited paper, Elliott, Ackerman, and Millian argued more than twenty years ago that a flurry of state regulatory activity often spurs a federal response as industry clamors for centralized regulation. They claimed that a high degree of state environmental regulatory activity can spur uniform federal legislation as a result of pressure from the regulated community. While this dynamic may, to be sure, explain some developments in environmental law, it is clearly not a satisfying explanation for state climate change action to date. Instead, my claim reverses theirs: federal law has spurred state regulatory activity by bolstering state regulatory capacity and leadership, leading ultimately to climate change regulation.
Iterative federalism schemes also shed light on the ongoing debate about devolution versus centralization in environmental policy making. In an influential article, Revesz argued that states are more likely to produce efficient levels of environmental regulation because of interstate competition for capital and residents.²³ The article led to a robust academic debate about federalism and environmental law, focused to a large extent on which level of government—state or national—will provide the optimal level of environmental services. Proponents of state devolution base their preference for state regulation principally on Tieboutian-influenced economic models about interstate competition, which predict that states will compete among themselves to produce an efficient level of regulation.²⁴ Centralization proponents, by contrast, argue that the nationalization of environmental law overcomes various market failures, including lax environmental standards among states that “race to the bottom” in an attempt to attract business; economies of scale in federal regulation; and controlling interstate externalities.

A close examination of iterative federalism schemes suggests that innovative regulatory mechanisms can have their cake and eat it too. These schemes simultaneously permit some of the chief benefits of devolution—policy experimentation, avoidance of untested and potentially expensive national mandates—while addressing interstate externalities, national product market economies of scale and the race to the bottom. These iterative federalism schemes also test empirically the contrasting hypotheses about devolution and centralization. For example, California’s experience in regulating mobile sources bolsters claims of centralization proponents that regulators often operate under conditions of scientific uncertainty and with poor information about the economic effects of their regulatory proposals. This example thus offers illustrative evidence suggesting that claims about a working competitive regulatory market among states are overstated. But these examples challenge the pro-centralization camp’s assumptions as well, for California’s experience demonstrates a significant benefit of devolution: minimizing the risk of overly stringent national regulation while allowing individual states to experiment and take risks. Premature federal adoption of California’s stringent emissions standards might have proven much costlier than allowing California first to experiment and then to have federal standards develop out of the California experience. Similarly, the experience with the Ozone Transport Commission—which adopted a ten state regional cap and trade scheme to regulate nitrous oxides (NOx)—provided an experiential base to use in persuading the federal government to expand the program’s reach to areas of the country much less politically supportive and to overcome potential public choice pathologies at the federal level. By the same token, the OTC states were pushed to develop stringent NOx-reducing
strategies by their need to comply with national air standards, standards that form the lynchpin of the centralized federal role in controlling air pollution.

The iterative federalism schemes analyzed here raise interesting possibilities for other pollution problems and for regulatory experimentation outside the environmental arena. Federal preemption, for example, has occurred in numerous substantive areas in recent years—including securities regulation; pension benefits; predatory lending; cigarette labeling and advertising; tort law; and liability for oil spills,\textsuperscript{25} often at the behest of industry.\textsuperscript{26} Though the case for uniform national standards in product markets has some intuitive appeal, one can imagine iterative federalism schemes in various substantive areas in which a particular state or states might be singled out to continue to play a regulatory leadership role, as California has, while preempting other states from regulating in order to avoid the chaos of fifty separate regulatory schemes. In the environmental arena, for example, all fifty states are preempted from setting energy efficiency standards for many appliances. Why not provide super-regulator status for California and let the state experiment with tighter standards? Similarly, regional problems like the management and transport of waste, water pollution, and traffic and land use might benefit from the regional approach embodied in the OTC, with strong state involvement bolstered by significant technical and leadership support from the federal government. In short, iterative federalism ought to expand our regulatory horizons.

2 See, e.g., BARRY G. RABE, PEW CENTER ON GLOBAL CLIMATE CHANGE, GREENHOUSE & STATEHOUSE: THE EVOLVING STATE GOVERNMENT ROLE IN CLIMATE CHANGE (2002); Kirsten H. Engel, Harnessing the Benefits of Dynamic Federalism in Environmental Law, 56 EMORY L.J. 159 (2006) (“Despite exhortations to the contrary, the federal government continues to address issues of purely local effect while the states continue to address issues of national—and even international—effect”); PEW CENTER ON GLOBAL CLIMATE CHANGE, CLIMATE CHANGE 101: STATE ACTION, http://www.pewclimate.org/docUploads/101_States.pdf (last visited June 10, 2008) [hereinafter CLIMATE CHANGE 101] (“In the absence of federal leadership to address climate change, many states and regions have begun taking action on their own.”).

3 See CLIMATE CHANGE 101, supra note 2, at 4-5; Barry Rabe, Race to the Top: The Expanding Role of U.S. State Renewable Portfolio Standards, 7 SUSTAINABLE DEV. L. & POL’Y 10 (2007). I use the word “surprising” because as many observers have noted, state climate change legislation presents something of an intellectual puzzle. Climate change is a classic global commons problem—any one jurisdiction acting alone cannot solve the climate problem by reducing greenhouse gas emissions so the incentives to continue polluting outweigh any incentives to reduce greenhouse gas emissions. See, e.g., Kirsten H. Engel, Mitigating Global Climate Change in the United States: A Regional Approach, 14 N.Y.U. ENVTL. L.J. 54, 55 (2005).

4 California, for example, has passed legislation to reduce its carbon emissions to 1990 levels by 2020, see CAL. HEALTH & SAFETY CODE §§ 38500 et seq. (West 2007), and the governor through executive order has called for emissions reductions of 80 percent below 1990 levels by 2050, see Executive Department, State of California, Exec. Order S-3-05.


7 I do not mean to suggest that areas in which state and federal policy making jurisdiction overlap cannot produce a dynamic, iterative process of federalism. Indeed many federal policies have been influenced heavily by earlier state policy iterations, including in the environmental area. California’s leadership on energy efficiency standards is an example. Instead, I suggest that the particular regulatory relationship embodied in the iterative federalism schemes I analyze produces a particularly robust iterative relationship that more self-consciously promotes an iterative, back-and-forth process of federalism.


10 See Regional Greenhouse Gas Initiative, supra note 6.

11 For a listing of the current NAAQS see ENVTL. PROT. AGENCY, DRAFT REPORT ON THE ENV’T 2003 1-6 (2003) [hereinafter DRAFT REPORT]. The human health standards are designated primary standards and the physical environment standards (including crops, ecosystems, and visibility) are called secondary standards. Id.

12 42 U.S.C. §§ 7409, 7410 (2000) (CAA Sections 109, 110). The Act contains important exceptions to cooperative federalism. For example, the federal government regulates mobile source emissions almost exclusively, 42 U.S.C. 7543(a) (2000) (though California has special authority to enact standards more stringent than the federal standards, see 42 U.S.C. § 7543(b) (2000)).


14 DRAFT REPORT, supra note 11, at 1-9-1-10. Non-attainment does not, however, signal failure in the war against air pollution. To the contrary, air quality has improved significantly since the 1970s.


18 ANDREW AULISI ET AL., WORLD RESOURCES INST., GREENHOUSE GAS EMISSIONS TRADING IN U.S. STATES: OBSERVATIONS AND LESSONS FROM THE OTC NOX BUDGET PROGRAM 11 (2005), available at http://www.wri.org/publication/greenhouse-gas-emissions-trading-us-lessons-from-otc-nox. Though these emissions reductions are impressive, it is unclear to what degree NOx emissions would have declined by similar amounts had each state individually regulated sources using more traditional command and control measures.


20 Id. at 13-14. The authors measured leakage in part by measuring changes in electricity supply and demand in the region, finding that net imports of electricity changed little in the pre- and post-trading program periods.


22 See E. Donald Elliott, Bruce A. Ackerman & John C. Millian, Toward a Theory of Statutory Evolution: The Federalization of Environmental Law, 1 J.L. ECON. & ORG. 313, 327 (1985) (noting that environmental issues often become federalized at the behest of industry groups who seek to avoid active and potentially disparate state regulatory activity).


26 See Hills, supra note 25, at 19.
Doug Lichtman joined the faculty at UCLA in 2007, after a tenured teaching career at the University of Chicago. His areas of specialty include patent and copyright law; telecommunications regulation; and general legal issues pertaining to technology firms and the Internet. Lichtman’s academic work has been featured in journals including the Journal of Law & Economics, the Yale Law Review, and the Harvard Business Review. He also writes for mainstream publications like the Wall Street Journal and the Los Angeles Times. In addition to his academic pursuits, Mr. Lichtman maintains an active consulting practice, advising Fortune 100 clients on patent and copyright strategy. He is currently serving as Scholar in Residence at the Los Angeles office of Loeb & Loeb. Mr. Lichtman has undergraduate degrees in electrical engineering and computer science from Duke University, and a J.D. from Yale Law School.
GOOGLE BOOK SEARCH IS NOT FAIR USE

Douglas Lichtman

In more than a decade of writing legal scholarship, I have never before been tempted to write a paper focusing exclusively on a single case. Yet the pending copyright litigation over Google Book Search has prompted me to break my own unwritten rule. The reason, quite simply, is the undeniable allure of the accused infringement. Most copyright cases have an entirely different aura, featuring accused infringers whose actions are plainly selfish in nature. But the Google Book Search project promises this amazing resource through which all of us would be able search the world’s books in much the same way that Google today allows us to search the Web. If courts and commentators can appropriately stand up for copyright even against such an alluring alleged infringer—if we can get the analysis right even when our first intuitions might be to ignore the law and simply cheer on Google in its widespread, unauthorized copying—then I think the copyright community will breathe a justified and important sigh of relief. After a decade where copyright law and its important role have been given short shrift in not only everyday life but also in much of the academic and legal commentary, the Google Book Search case stands simultaneously as both a chance and a challenge to finally and thoughtfully right the ship.

Google is in the process of creating an online search engine that allows users to search the full text of published books. To use the search engine, users enter a search term or phrase, and Google’s computers then look for books that might use the term or phrase and hence might be of interest. The books about which there is controversy are books that Google obtains from various libraries. The libraries allow Google to borrow books from their collections, scan those books into electronic form and ultimately include the resulting electronic information in whatever databases Google builds in order to run its search service. The libraries do not hold copyright in the books and thus the libraries themselves have no power (from a copyright perspective) to authorize Google’s use.

Google scans the books it borrows in their entirety, and Google stores all of that information in a way that allows Google to respond to any search query that might
be submitted in the future. Thus, presumably, Google saves all or most of the text of every book in some sort of database. Users of Google Book Search, however, do not see the full text of a book unless the relevant copyright holder has given permission. Instead, Google returns what it describes as “snippets,” which seem to be excerpts that run only a few sentences long and contain the desired search terms. These excerpts in theory show the user enough information that the user can evaluate whether a given book is indeed of interest. Google has proprietary software that is designed to ensure that users cannot see too many excerpts from the same book, for example through repeated searching.

Google has publicly committed to leave certain books out of its database, including thesauruses and anthologies of short poems. Google unilaterally decides which books to leave out, but the idea is to exclude books where most of the value of the book comes from having the ability to access a small relevant excerpt at the right time. Google has not published a list of the books excluded, nor has it made public the details of how it selects these titles. Google also allows copyright holders to “opt out” of the Google Book Search program. Specifically, a copyright holder can notify Google that it would prefer to have a specific work removed from the database. Google presumably complies with these requests.

There are a number of services that compete with the Google service. Amazon, for example, has implemented and announced a variety of search-inside-the-book programs, including a voluntary program through which copyright holders can allow would-be customers to “look inside” a book prior to buying it, and an announced program that would (among other things) allow users to electronically search participating books after they have purchased the relevant book in paper form. The book publisher HarperCollins is also experimenting with electronic delivery. And even Google itself has launched a competing service—one that waits for permission from copyright holders, but upon receiving permission reports back larger excerpts. Many other services and products are similarly either available today or in various stages of negotiation and development.

II. The Case

Litigation is already underway over the Google Book Search project. The result of that case will ultimately turn on the court’s interpretation of section 107 of the Copyright Act. Section 107 empowers a court to excuse, on public policy grounds, acts that would otherwise be deemed to impermissibly infringe a copyright holder’s exclusive rights.
Courts are required to consider four specific statutory factors when evaluating a fair use claim; however, courts are empowered to go beyond those factors and engage in a broader public policy analysis as appropriate.\(^4\) In the end, the idea is for courts to excuse infringement in instances where a “rigid application of the copyright statute… would stifle the very creativity which that law is designed to foster.”\(^5\) The fair use doctrine is thus enormously flexible, and by necessity it vests considerable discretion in each court.

A common misconception is that the fair use doctrine excuses any infringing use that is socially valuable. That is a clear mistake. *Princeton University Press v. Michigan Document Services* provides a helpful example.\(^6\) The infringing products in that dispute were packets of photocopied materials. The packets were made up of excerpts from articles and books, those excerpts having been chosen by university professors for use in their specific university classes. The accused infringer was the copy center that duplicated the excerpts and ultimately sold those packets to students.

Clearly, the infringing products were socially attractive. They were products that facilitated classroom teaching, and they were produced at the direction of university faculty. Yet, the copy center that produced the packets was found guilty of copyright infringement and specifically had its fair use defense rejected.\(^7\)

Why was the copy center denied the protection of the fair use doctrine? Because fair use is not an inquiry into whether the accused use is valuable. Instead, it is an inquiry into whether the owner of the infringed copyright should have influence over when and how the accused use takes place. To deny fair use in *Michigan Document Services*, then, was not to in any way speak ill of the infringing products at issue. Photocopied university materials are tremendously worthwhile products, and no one disputes that fact. To deny fair use was instead to decide that these beneficial but infringing products ought to fall under copyright holders’ sphere of influence, with the relevant copyright holders having the right to influence who produces the packets, under what terms and how much everyone profits from that interaction.\(^8\)

Two intuitive considerations guided the court in *Michigan Document Services* and indeed more generally seem to helpfully frame fair use analysis. The first of these intuitive considerations is the degree to which a finding of fair use would undermine the incentives copyright law endeavors to create. Copyright law in general recognizes rights in authors in order to motivate authors to create, disseminate and in other ways develop their work.\(^9\) Fair use is unattractive to the extent that it interferes with that goal.
Put differently, the issue here is whether repeated findings of fair use in a category like the one at issue would over the long run reduce author motivation to do things like create their work, share their work publicly and search for new related projects.\textsuperscript{10} If so, fair use is on this ground unattractive, as it undermines the very incentives copyright law endeavors to create.

The second intuitive consideration relevant to fair use analysis is the degree to which uses like the one at issue could survive without the protection of fair use. In \textit{Michigan Document Services}, for instance, there was little doubt that university reproduction would continue regardless of whether fair use was recognized. With fair use, reproduction would take place under the combined control of the copy center and faculty member. Without fair use, copyright holders would for the most part license this use, anxious to earn the additional royalties associated with classroom adoption and cognizant of the fact that a faculty member can always assign other reading if a given copyright holder asks for an unreasonable price or imposes unreasonable terms. Either way, then, course materials would be created.\textsuperscript{11}

Contrast that example with an example involving a classic fair use, parody. A parody is a work that borrows from some preexisting work in order to poke fun at or in other ways critically comment on the original.\textsuperscript{12} Copyright holders might refuse to authorize parodies in a world where permission is required. Parodies are thus an attractive candidate for fair use because fair use might be the only practical way to ensure that society gets them.\textsuperscript{13}

Return now to Google Book Search. To the extent that Google invokes fair use to defend the entire Google Book Search program, that defense in my view fails. A finding that Google Book Search is fair use would clearly hurt authors. For instance, Google’s scanning and storage activities expose authors to an increased risk that their works will leak out in pirated form. And Google’s project more generally undermines an author’s incentive to implement and profit from comparable or competing offerings. Moreover, a finding of fair use is not critical in terms of facilitating the creation of the book search engine, because a great deal of the project could be accomplished through negotiated, consensual transactions.

Were Google to concede infringement for many of the works at issue but invoke fair use only to more narrowly excuse its use of books in instances where the costs of identifying the relevant copyright holder is prohibitive, Google’s claim would be strong. It is enormously difficult to acquire permission with respect to books that
are significantly old, or books for which the current ownership of rights is hopelessly unclear. As applied to that narrow class, Google might be right that the only way to use those books is to invoke fair use.\textsuperscript{14} Google could also fairly point out that the harm to that subclass of authors is small, because authors who are so difficult to identify are likely also not authors who are actively profiting from or marketing their work. The main weakness with this argument is that Google in practice makes no effort to distinguish these “orphan” works from the many works for which permission would be practical. A court might require Google to undertake reasonable efforts along these lines as a condition of any fair use finding.

In summary, then, Google's fair use claim fails, in my view, because Google's legal argument and its actual practices both sweep too broadly. Google has a narrow but strong claim with respect to certain works that it includes in its search database. But that narrow claim does not immunize the project more generally.

\textbf{Fair use is an affirmative defense to a charge of copyright infringement. Its purpose is to “permit[ ] courts to avoid rigid application of the copyright statute when…it would stifle the very creativity which that law is designed to foster”\textsuperscript{15}} Fair use began as a flexible, judge-made doctrine. When federal copyright law was revised in 1976, however, fair use was codified in the statute at section 107. That codification was explicitly intended to re-state the then-existing law and not to expand or contract fair use in any way.\textsuperscript{16} Thus, even today, fair use retains the flexibility and comprehensiveness of an equitable doctrine.

The statutory provision that codifies fair use begins with a list of examples, stating specifically that “reproduction…for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research” is excused.\textsuperscript{17} The provision then goes on to identify four factors that must be considered when evaluating a claim of fair use. Those factors are:

- the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- the nature of the copyrighted work;
- the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- the effect of the use upon the potential market for or value of the copyrighted work.\textsuperscript{18}
These factors are not exhaustive. Thus, courts can and do consider other factors when conducting a fair use inquiry, emphasizing facts that might not fit within the normal rubric but still seem important to understand the dispute at hand. Moreover, when considering the four explicit factors, courts do not merely count them up. Instead, courts combine these factors with other relevant information and conduct an appropriately flexible, case-specific policy analysis. “The ultimate test of fair use...is whether the copyright law’s goal of promoting the Progress of Science and useful Arts would be better served by allowing the use than by preventing it.”

Courts typically organize their fair use analysis by first considering each of the statutory factors and then, as needed, turning to other considerations. I adopt that same framework here and discuss each of the four statutory factors, apply them to the facts at hand, and then consider issues that do not fit well under those four headings.

III. A. The Purpose and Character of the Use

The first fair use factor is the purpose and character of the use. One issue typically raised with respect to this factor is whether the use is commercial. The intuition is that a profit-generating user can, and thus should, absorb the costs of complying with copyright law and compensating the original author.

There was a time when this consideration was significantly influential. In Sony v. Universal City Studios, for instance, the Supreme Court stated that “every commercial use of copyrighted material is presumptively an unfair exploitation of the monopoly privilege that belongs to the owner of the copyright.” More recently, however, the Court has backed away from this strong stance, holding instead that “the commercial or nonprofit educational nature of a work is not conclusive” and is only one factor “to be weighed along with others in fair use decisions.”

The reason for this hesitation is simple: many commercial uses are at the same time strong candidates for fair use. Newspapers and television stations, for instance, are clearly for-profit entities engaged in for-profit uses. Yet, to the extent they commit copyright infringement, they typically do so in support of the news reporting and commentary functions that are explicitly endorsed in section 107. The fact that an entity has a profit motive, then, turns out to not be particularly helpful in terms of distinguishing attractive from unattractive fair use cases. At best, the commercial nature of a use serves as a weak signal that the infringer has resources that could be used to reward or empower the original copyright holder, and that a requirement to do so would not substantially reduce the availability of the work in question.
A second and more important issue considered as part of the first factor is the question of whether the accused use is “transformative” in nature. A use is transformative if it is substantially different from the original work in terms of its purpose, meaning, or effect.26 A transformative work does not merely supersede the original work. It is instead a work that has new features or brings new value.27

Whether a work is transformative is important for two reasons. First, all else held equal, a transformative work is less likely to hurt the original author. If an infringing work has the same purpose, meaning or effect as does the original work, the infringing work likely will displace sales of the original. If the infringing work is sharply different along these dimensions, by contrast, sales could remain intact.28

The second reason why it is important to consider whether a work is transformative is that a transformative work brings something valuable to society. The work is not merely redundant to that which society already had. It is new and has new meaning. The fact that a work is transformative, then, makes a finding of fair use marginally more attractive. Put differently, there is little reason to trump a copyright holder’s exclusive rights if the only payoff is that society would get another work that is largely indistinguishable from the original one. By contrast, if society is at least getting something sufficiently new, there might be a case for a fair use finding, because getting something new is itself an attractive outcome.29

Applying all this to the Google Book Search project, the commercial nature of the use is straightforward: Google clearly is a for-profit entity engaged in a profit-motivated use designed to promote Google’s long-run financial interest. Indeed, if Google were spending this much money and not anticipating an ultimate return on the investment, Google’s management team would likely be violating its fiduciary duty to Google’s stockholders. The fact that Google is not at the moment explicitly cashing in on the infringing product is of little importance. Clearly, over the long run, Google will monetize its new search engine, perhaps by introducing advertisements, or by demanding a royalty on downstream book sales, or by using this new search capability to further distinguish the Google family of products from rival products offered by firms like Microsoft and Yahoo.30

With respect to the transformative nature of the work, however, Google has a strong case that Google Book Search is transformative. The overall purpose of Google’s infringement is to create a new and useful tool for locating information. I do not think that tells us much about whether a finding of fair use hurts author incentives. But
it does tell us that there is at least something to be gained by a finding of fair use. Google would, if protected by fair use, put into the world a product that is both socially valuable and meaningfully distinct from the works that are being infringed. In my view, that suffices to establish that the use is transformative.31

The second explicit fair use factor is the nature of the copyrighted work in question. Under this factor, courts consider the creativity of the original work. If the original work falls into a highly creative category, such as fictional novels, fair use is deemed less compelling. If the original work falls on the less creative side of the spectrum, such as a biography, fair use is deemed more appropriate. The explanation is that “some works are closer to the core of intended copyright protection than others.”32 Put differently, on this view, copyright law is primarily concerned with the protection of creative, expressive work, and as a result fair use is less objectionable when it reduces the protection given to works that are not significantly creative or expressive.

A second consideration sometimes included in a discussion of this fair use factor is the question of whether the original work is sufficiently available to the public. A work that is out of print, for example, might on this argument be more vulnerable to a fair use defense.33 The intuition here is two-fold: first, fair use might be the only way to facilitate use of an otherwise unavailable work; and, second, a finding of fair use might not much undermine author incentives in a situation where the author has himself already stopped promoting or otherwise offering his work to potential licensees.

Application of this unavailability concern is complicated, however, and courts have varied in their approach. In Basic Books v. Kinko’s Graphics, the court noted that it might be more important to deny fair use as applied to out-of-print works because the royalties at issue in the litigation “may be the only income” the relevant authors will earn.34 In Princeton University Press v. Michigan Document Services and separately in American Geophysical Union v. Texaco, two courts recognized that, by denying fair use, copyright law can support the development of intermediaries like the Copyright Clearance Center that facilitate licensing and in that way make more work accessible.35 The influential Nimmer treatise, meanwhile, makes a related point: an out-of-print work will come back into print whenever demand is high enough and costs are low enough, but those conditions might “never arise if competitors may freely copy the out-of-print work.”36

Applying all this to Google Book Search produces a mixed result. Some of the
infringement that takes place as part of the project would likely be favored under the second fair use factor, either because the books being infringed are more informational than creative, or because the books are out of print and/or otherwise inaccessible for licensing. However, to the extent that Google scans books that are largely creative, or to the extent that Google scans books that are in fact available for consensual licensing, the second fair use factor would likely favor the copyright holders.

Interestingly, note that Google does not separate books along these dimensions when it engages in its infringing activities. It could. Google’s partner libraries surely sort their collections in ways that distinguish novels from biographies. And it would be easy for Google to check, prior to scanning, whether a given book is in print or is otherwise available for licensing through its author, publisher, or a licensing intermediary. This failure on Google’s part might be deemed to forfeit Google’s otherwise legitimate claim to a partial victory under factor two.

The third explicit fair use factor is the amount and substantiality of the portion used. As a general rule, the more the infringer takes, the more this factor weighs against a finding of fair use. The intuition is the obvious one: the extent of the copying is a good proxy for the harm imposed on the copyright holder. If an infringer takes only a tiny segment of a copyrighted work, the odds are low that the taking will much undermine the author’s ability to exploit his own full contribution. If the infringer takes the bulk of the work, the opposite logic applies. In this sense, this third factor in some ways echoes the considerations raised under the first factor’s test for transformative use and the fourth factor’s test for the economic significance of the copying.

There are exceptions to the general rule stated above. For instance, copying a small amount from the original work might still be problematic under this factor if what was taken turns out to be “essentially the heart” of the work.37 Conversely, copying the entire work might not weigh against fair use in a case where the only way to accomplish the infringing use is to copy at that scale.38

The Google Book Search project obviously involves the scanning of entire books, and thus to some degree the third factor will weigh against a finding of fair use. This is appropriate because it is the existence of these full copies that leads to one of the harms that most concerns copyright holders: full copies might accidentally leak out. That distinguishes the aforementioned cases where copying of the full work was excused. In those cases, full copies were made, but there was never much risk that those full
copies would fall into the hands of unrelated parties. Here, the risk is significantly more pronounced.

Pushing in the opposite direction, however, note that while a workable search engine could be built through a process that used less than the full text of the relevant books, the charm of the Google project is that its search engine can search any word or phrase in the book. That is what makes Google’s search index better than conventional alternatives. There are many indexes that sort books based on keywords or other organization themes that are chosen ahead of time by the organizing party. Google’s index is unique in that it allows the user to dynamically define the keywords that will then be used to retroactively sort the books. That feature could not be achieved without Google having access to the full text of the works.39

Putting all of that together, I doubt that the third factor should or will much move a court’s analysis one way or the other. As I suggest above, the third factor is largely redundant to the analysis conducted under the first and fourth statutory factors. I suspect that the third factor will, therefore, not be paid much attention. The other two are in this case much more helpful in terms of sharpening the core public policy issues at stake.

The fourth explicit factor listed in section 107 is the effect on the potential market for, or value of, the copyrighted work. This is relevant because a use that interferes with the value of the original work likely undermines the incentives that copyright law is designed to create in the first place. That is, the whole idea behind copyright law is to encourage authors to create, disseminate, and in other ways promote their work by promising authors certain exclusive rights. The more a fair use finding would reduce the value of those exclusive rights, the more disruptive that fair use is to the copyright system, and hence the less attractive the fair use defense.

When evaluating the fourth factor, courts consider “not only the extent of market harm caused by the particular actions of the alleged infringer, but also whether unrestricted and widespread conduct of the sort engaged in by the defendant would result in a substantially adverse impact on the potential market for the original.”40 That is, the fourth factor does not merely look to see whether this infringer would, through its actions alone, substantially impose author harm. The factor more broadly considers whether actions in this category, if repeated by a large number of unrelated infringers, would cause substantial author harm.
That harm, meanwhile, includes harm to “potential” markets. Thus the fourth factor is implicated not merely when the infringing use might reduce sales of the original work in its current form, but more generally when the infringing use might interfere with future exploitation of the work in other forms.\textsuperscript{41} Relevant markets under the fourth fair use factor include markets that the author has not yet entered.\textsuperscript{42} One influential line of cases holds that any market can count as long as it is a “traditional, reasonable, or likely to be developed” market.\textsuperscript{43}

Courts and scholars sometimes worry that this fourth factor is circular.\textsuperscript{44} After all, if fair use is denied in a given case, then the infringer in that case would himself likely pay the author some sum in exchange for the right to continue the infringement. Can that potential payment really count under factor four, the result being that in almost every case factor four would, at least to a small degree, weigh against a finding of fair use?\textsuperscript{45}

The answer is that factor four actually should in every dispute weigh at least slightly against a finding of fair use. This is not to say that fair use should be denied in every case. Instead, my point is that, in almost every case, fair use does reduce author incentives. Other considerations might then swamp that concern; but factor four is designed to highlight the degree to which a finding of fair use would hurt authors, and framed that way there is no reason to exclude from the calculus the losses associated with the very use being litigated.\textsuperscript{46}

On the facts of Google Book Search, the fourth factor weighs strongly against a finding of fair use because there are at least four types of cognizable harm.

First and most obviously, Google imposes a substantial harm on authors when it scans, transmits and stores complete electronic copies of previously non-electronic books. The harm here comes in the form of a security risk. Google’s electronic copies could leak out not only during the initial scanning process but also later in time, when the electronic copies are stored indefinitely in Internet-accessible databases. Google surely has security precautions in place to prevent the electronic versions from leaking out. However, there is no reason to believe that Google’s security precautions are appropriate from a copyright holder’s perspective.\textsuperscript{47} Put differently, copyright holders are harmed here because electronic duplication introduces new and substantial risks, and yet Google’s project allows copyright holders no say over how those risks should be managed or what should happen in the event the risks mature into a substantial security breach.
Google would likely respond by suggesting that courts can evaluate Google’s security precautions and make any fair use finding contingent upon a showing of adequate security. That is in part an attractive middle ground. Factor four analysis, however, cautions against that approach. After all, the question here is whether “unrestricted and widespread conduct of the sort engaged in by the defendant” would impose substantial author harm. In my view, if authors are told that anyone can scan, transmit and store full copies of their books for use in index-like products, and that the only protection is after-the-fact judicial evaluation of the relevant infringer’s security precautions, I suspect that authors will rightly expect that their work will leak out. Courts are just too slow and too far removed from technical details to meaningfully regulate security issues of the sort implicated here.

Second, for at least some of the works being copied, Google’s act of providing snippet access will directly undermine the market for the original works. A technical dictionary, a thesaurus, an anthology of short poems and a book of famous quotations are each valuable in large part because users are at any given time interested in only a specific short snippet excerpt. If Google provides those very excerpts via its online search engine, the value of these books will be sharply reduced.

As I mentioned earlier, Google itself has acknowledged this and made a public commitment not to provide even snippet access to these sorts of works. As with the security issue, however, that solution is unsatisfying both because Google’s judgment might not align with authors’ judgment, and because again the proper analysis here is to consider not merely whether authors would be harmed if forced to trust Google on this matter but more generally whether authors would be harmed if snippet access of this sort were to become a widespread practice, run by possibly trustworthy firms like Google but also by a wide range of actors with varying degrees of honorable motivation.

Third, Google’s project directly undermines author opportunities to pursue projects that are similar to and/or partially competitive with Google Book Search. For instance, both Amazon and the publisher HarperCollins have announced their own services that would include electronic book access and/or book search capabilities. If Google is allowed to compete with those services under the protection of fair use, authors will have a harder time earning profits from and otherwise being successful with these other programs.

Fourth and finally, there is the purely circular harm: if Google’s fair use defense is
rejected, Google will surely take steps to include authors in the design of the book search project and also to include authors in some of the financial gains the service makes possible. As I note above, this circular harm is a controversial consideration, but in my view the circular harm is rightly included in the factor four calculus. Again, the question under factor four is the degree to which a finding of fair use would limit author control and author profit, thereby undermining author incentives. Google’s refusal to include authors in the decision-making process and its decision to deny authors any share of Google’s revenues is therefore plainly relevant. If Google Book Search is even half as successful and socially important as its proponents predict, the royalties at issue in this case alone could significantly increase author incentives to write, disseminate and otherwise invest in their work.

The four statutory factors play a central role in almost any fair use analysis. However, fair use also welcomes consideration of other relevant public policy issues. Here, then, I briefly consider two issues that the parties might raise along these lines.

The popular commentary on Google Book Search emphasizes the fact that Google’s search engine will likely increase demand for books. That argument resonates. By making it easier for people to identify books that might be of interest, a comprehensive search engine should in the aggregate increase book demand. This should be especially true for books that serve a niche market. Those books are hard to find in conventional ways because they are not sufficiently known or advertised, but Google’s content-based search engine should compensate for those limitations, increasing the likelihood that interested readers will find these niche offerings.

That said, the fact that the Google project might in one way benefit copyright holders does not significantly change the overall fair use analysis. After all, this fact tells us only that authors are better off in a world where Google’s project is fair use as compared to a world where no one builds book search engines at all. That, however, is not the relevant comparison. Instead, the fourth factor of the fair use inquiry asks about the degree to which authors are worse off in a world where fair use takes away their ability to license the use or pursue it themselves. Clearly, authors would be better off if they could negotiate their own deal with Google, or pursue their own versions of the search technology, rather than merely receiving whatever sales benefit the project happens to offer them by default.
None of this should be surprising. All sorts of infringing work benefits authors, and yet authors nevertheless routinely keep their right to say no. Movies that are based on books, for example, typically increase demand for the underlying books. Still, there is no question that the people who produce those movies must ask permission from, and negotiate financial details with, the relevant copyright holders. The reasons are the very ones I have considered at length here: author incentives are at stake in the question of whether or not a movie should fall under the copyright holder’s sphere of influence; and, because movies really do create value that can be shared by both the movie producer and the relevant book author, it seems likely that movies will still be made even if fair use is denied.

The popular commentary also has been taken with the argument that Google’s use should be deemed fair because Google allows copyright holders to opt out of the program. Specifically, the relevant copyright holder can notify Google that he does not want a particular book included in the database, and Google has promised to respect that request.

This opt-out offer certainly makes the Google project more attractive than it would otherwise be, but again my suspicion is that this feature will not and should not significantly influence the overall analysis. The reason is the fundamental insight that fair use considers “not only the extent of market harm caused by the particular actions of the alleged infringer, but also whether unrestricted and widespread conduct of the sort engaged in by the defendant would result in a substantially adverse impact on the potential market for the original.” An opt-out works well in a world where Google is the only infringer. Authors could in that case at low cost find out about the Google project and communicate their desire to be left out if need be. This would be efficient, in fact, because the costs to authors in finding Google would likely be much smaller than the costs Google would incur were it required to find each individual copyright holder.

When the analysis shifts to focus on the possibility of countless Google-like opt-out programs, however, the conclusions reverse. In a world with a large and ever-changing list of opt-out programs, authors would be forced to invest substantial sums finding each opt-out program and notifying each about their desire to participate. The problem would be even worse if some of those opt-out programs were designed strategically to make things difficult on authors, for instance imposing high standards of proof before acknowledging that an opt-out really came from the correct copyright holder.
(Infringers have an incentive to do just that, because in an opt-out system infringers benefit if authors find it too expensive to actually engage in the mechanism of opting out.)

Overall, then, the problem with opt-out is that it does not scale. This is one reason why copyright more generally is defined as a permission-based, opt-in system. The opt-in approach gives copyright holders meaningful control over potential infringements. Opt-out, by contrast, is an expensive proposition that would substantially erode the value of copyright rights.49

As I have emphasized repeatedly, fair use analysis is inherently subjective. My own view is that the fair use defense should be and will be rejected in the context of the Google Book Search project. My goal here, however, has been to explain the underpinnings of my position, so that my rationales and understandings can be compared against analysis put forward by others who might to varying degrees disagree with my conclusions.

Two points warrant final emphasis. First, on the side of the copyright holders, the most important point is that Google’s project really does undermine the long-term value of their work. Copyright holders will find it difficult to control and profit from similar projects if Google is allowed to pursue its project without their permission; and there is no reason to cripple authors’ future in that way. After all, even without fair use, Google or some similar firm will be able to build exactly this sort of useful tool. Permission is not a death knell for innovation. It is instead a way to make sure the copyright rights remain meaningful even as technologies and needs change.

Second, on Google’s side, the most important response is that for certain works permission is impractical. The copyrighted work might just be too old, or the contracts that originally allocated the copyright rights might today be lost or impenetrable. Obviously, where permission is not practical, a legal rule requiring permission is unwise. The result would be to functionally bar the downstream use, and to do so in a setting where there likely is no active copyright holder ready to benefit from that extra control. Admittedly, a court might reasonably require that Google take steps to distinguish orphan works from works that are more actively being cared for. But, that issue to one side, fair use is attractive as applied to genuine orphan works.
This article is excerpted from a larger, in-progress treatment of the fair use doctrine. Comments welcome at lichtman@law.ucla.edu.

See Press Release, Authors Guild, Authors Guild Sues Google, Citing “Massive Copyright Infringement” (Sept. 20, 2005).


Stewart, 495 U.S. at 236 (quoting Iowa State Univ. Research Found., Inc. v. Am. Broad. Cos., 621 F.2d 57, 60 (2d Cir. 1980)).


Id. at 1385-90 (discussing fair use).

See, e.g., id. at 1387 (discussing how a permission-based copyshop system would work).

Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) (“The immediate effect of our copyright law is to secure a fair return for an author’s creative labor. But the ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good.” (citations omitted)).


See Michigan Document Servs., 99 F.3d at 1387 (discussing permission fees); Am. Geophysical Union v. Texaco Inc., 802 F. Supp. 1, 7 (S.D.N.Y. 1992) (discussing the Copyright Clearance Center as an example of an intermediary that helps entities like copycenters clear necessary permissions).


Id. at 1178 (emphasizing the “unlikelihood that creators of imaginative works will license critical reviews or lampoons of their own productions”). There are obviously other reasons why a given use might not adequately flourish in the absence of a fair use defense. For instance, it might be that the costs of negotiating permission are so high that a given infringing use will not be undertaken even though the relevant copyright holders would agree to permit the use if asked. Similarly, there might be some market failure in the market for the infringing good such that demand would be inefficiently low. In that case, the infringer would not be willing to pay enough for his use, and thus the use might not take place at appropriate levels in a permission-based system. My point in the text is therefore a broad one: fair use analysis ought to consider whether a denial of fair use might lead to an inefficient undersupply of the infringing work, no matter whether that undersupply is caused by copyright holders who decide not to license or some other economic or social factor.
My hesitation in this sentence comes only because it is easy to imagine the creation of a rights clearinghouse that would facilitate licensing of even these hard-to-license works. Indeed, enormous social value would be created were such a clearinghouse to be established, because that clearinghouse could then facilitate all sorts of uses of these works above and beyond the index use that Google is here litigating. For now, however, such a clearinghouse does not exist. It would therefore be relevant to a court's analysis only if the court believed that, by denying fair use in this case, the court could increase the likelihood that such a clearinghouse would come into existence.

*Campbell, 510 U.S. at 577* (citations omitted).


*Id.*

*See Castle Rock Entm't v. Carol Publ'g Group, Inc., 150 F.3d 132, 141 (2d Cir. 1998) ("[T]he four listed statutory factors in §107 guide but do not control our fair use analysis and are to be explored, and the results weighed together, in light of the purposes of copyright" (citations and quotations omitted))).

*See, e.g., Harper & Row Publishers, Inc., 471 U.S. at 560 (there is “no generally applicable definition [of fair use]” and “each case raising the question must be decided on its own facts” (quotations omitted)); Wright v. Warner Books, Inc., 953 F.2d 731, 740 (2d Cir. 1991) (“The fair use test remains a totality inquiry, tailored to the particular facts of each case.”).*

*Castle Rock Entm't*, 150 F.3d at 141 (citations and quotations omitted).

*See Itar-Tass Russian News Agency v. Russian Kurier, Inc., 886 F. Supp. 1120, 1129 (S.D.N.Y. 1995) (a “commercially exploitative use” is “one in which the copier directly and exclusively acquires conspicuous financial rewards from its use of the copyrighted material”); A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1015 (9th Cir. 2001) (even if the infringer does not actually profit from his bad act, “commercial use is demonstrated by a showing that repeated and exploitative unauthorized copies of copyrighted works were made [merely] to save the expense of purchasing authorized copies.”).*


*See Salinger v. Random House, Inc., 650 F. Supp. 413, 425 (S.D.N.Y. 1986) (“[P]ublishers of educational textbooks are as profit-motivated as publishers of scandal-mongering tabloid newspapers. And a serious scholar should not be despised and denied the law's protection because he hopes to earn a living through his scholarship.”).*
26 See Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 720 (9th Cir. 2007); Campbell, 510 U.S. at 579.

27 Campbell, 510 U.S. at 579.

28 Id.

29 See, e.g., Rogers v. Koons, 960 F.2d 301, 309 (2d Cir. 1992) (“The first factor…asks whether the original was copied in good faith to benefit the public or primarily for the commercial interests of the infringer”). One caveat to the above summary: some courts recognize a work as transformative only if the work is different from the original work in an expressive way. These courts do not accept evidence of just any new “purpose, meaning, or effect”; instead, they require a new expressive purpose, a new expressive meaning, or a new expressive effect. The rationale is that copyright law itself is designed to encourage expressive outputs and indeed itself refuses to protect valuable non-expressive works like databases and directories. See Feist Publ’ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991). Some courts therefore think it appropriate to similarly distinguish expressive from non-expressive work under the first fair use factor. Specifically, these courts refuse to recognize as transformative a work for which the new contribution is informational, organizational, or in some other way valuable but not expressive. See, e.g., Infinity Broad. Corp. v. Kirkwood, 150 F.3d 104 (2d Cir. 1998) (concluding that retransmission of radio broadcast over telephone lines is not transformative despite the fact that the retransmission was for an entirely different, albeit non-expressive, purpose).

30 Although, again, some courts will disagree, objecting that Google’s use might be valuable, but it is not expressive. See, e.g., Roy Export Co. v. Columbia Broad. Sys. Inc., 503 F. Supp. 1137 (S.D.N.Y. 1980), aff’d, 672 F.2d 1095 (2d Cir. 1982) (broadcast of unsponsored television program is still commercial because the broadcaster “stood to gain at least indirect commercial benefit from the ratings boost which it had reason to hope would…result” from airing the infringing program).

31 See Perfect 10, Inc., 487 F.3d at 720-21 (search engine’s infringement of copyrighted images found to be a transformative use despite informational nature of the use); Kelly v. Arriba Soft Corp., 336 F.3d 811, 818 (9th Cir. 2003) (same).

32 Campbell, 510 U.S. at 586.


37 Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566 (1985). In a famous case along these lines, a magazine purloined a tiny portion of an unpublished manuscript, but still the third fair use factor was deemed to favor the copyright holder because the copied words represented the excerpts that would-be readers were likely most interested in seeing. Id. (“The portions actually quoted were selected...as among the most powerful passages in those chapters.”).

38 In Sega Enter. Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992), for instance, the infringer copied the entirety of a software program in order to study how certain aspects worked. The court put “very little weight” on the amount of copying, however, both because the complete copy was not actually used after the learning was complete, and because there was no reasonable alternative means by which to dissect the program anyway. Id. at 1526-27.

39 The third fair use factor is similarly complicated as it applies to the snippets that Google offers to its users. Snippets in this context are in a very real sense the heart of each work. They are chosen based on the user's own search terms, and they are designed to show the user the exact part of the book that the user is most interested in seeing. Thus there is a very real analogy to be drawn to the Harper & Row case cited above. In both this case and that one, the size of the infringement is not a good proxy for its economic or artistic significance; the takings in both situations are small but tremendously well targeted.


41 Thus, in Campbell v. Acuff-Rose, the Supreme Court very nearly excused under the fair use doctrine a musical parody that happened to be written as a rap. The Court remanded to the lower court, however, for fear that the existence of a rap parody might significantly interfere with the original author's ability to license non-parodic rap versions in the future.

42 Campbell, 510 U.S. 569.

43 Am. Geophysical Union v. Texaco Inc., 60 F.3d 913, 930 (2d Cir. 1994).

44 See, e.g., id. at 930 n.17.

45 I say “in almost every case” rather than “in every case” because, in some cases, transaction costs would make it impossible for the accused infringer to pay even if the infringer wanted to.

46 See Am. Geophysical Union, 60 F. 3d at 931 (“The vice of circular reasoning arises only if the availability of payment is conclusive against fair use.”).

47 Discovery will reveal more information relevant to this discussion. For now, however, the already public contract between Google and the University of Michigan makes clear the mismatch between Google's incentives and author incentives. Google's contract imposes very few limitations on what Michigan does with the electronic copy of each book that Google provides to Michigan. Had the relevant copyright holders written the contract, surely
they would have more carefully articulated Michigan’s obligations to make sure that those electronic copies do not end up freely available on the Internet or in other ways abused.


49 It is possible to imagine that intermediaries would arise to search for new opt-out projects and opt out on behalf of participating copyright holders. That would reduce the overall costs of the opt-out approach, but it would be a complete waste from a social welfare perspective. In essence, the intermediaries would be re-creating the opt-in approach currently in place, but doing so in a more cumbersome and costly manner.
Professor Katherine Stone is a leading expert in labor and employment law in the United States. She was awarded a Guggenheim Fellowship Award in 2008 and a Russell Sage Fellowship for 2008–2009 for her work on the changing nature of employment and the regulatory implications. Her forthcoming book, *Globalization and Flexibilization: The Remaking of the Employment Relationship in the 21st Century*, will examine the changing employment landscape in Japan, Australia, and Europe. Professor Stone has been a member of the faculty of UCLA School of Law since 2004.

A LABOR LAW FOR THE DIGITAL ERA

Katherine V.W. Stone*

There is a serious problem with the labor and employment law system in the United States today. Unions have declined to the point where they represent less than eight percent of the private sector workforce, employee wages have stagnated for more than three decades, employers are cutting back on workers’ health insurance and pensions, and there is a dramatic growth in the numbers of the working poor. At the same time, there has been a rising chorus of complaints from labor scholars that the labor law has become “ossified,”¹ that the law is failing to offer meaningful worker protection,² that the courts and Labor Board have abandoned the “core values of labor law,”³ and that Congress has defunded the labor protective agencies such as the National Labor Relations Board (NLRB), the Occupational Safety and Health Administration (OSHA) and the Hour and Wage Division that administers the Fair Labor Standards Act (FLSA).⁴ Indeed, some have contended that during the past two decades, there has been a passive repeal of the employment statutes.⁵

There is a reason that the field of labor and employment law has declined. Work itself has not declined in importance—it remains a central part of individual identity and a prominent aspect of social life, but the labor and employment laws do not address the concerns or vulnerabilities of the majority of the workforce today. Instead, the regulatory framework governing the workplace is becoming irrelevant.

The question I address here is, how did the system come to this state and where is it going? In order to consider the future, one must develop an analytic and dynamic understanding of the present and the past. A future-oriented interpretation of the present and the past can help identify trends, provide a basis for critique and suggest constructive directions for change.

The American system of employment regulation is a two-track system. “Labor law” provides the mechanism for collective bargaining and other forms of employee collective action, while “employment law” sets minimal employment standards for all employees. Employment laws set minimum wages, establish safety and health standards, provide old age assistance, require unemployment insurance, compensate
industrial injuries, mandate child care and medical leave, and establish other minimal
terms of employment. Because the employment law standards are generally meant to
be floors, they do not obviate the need for workers to bargain, whether individually or
collectively, for employment standards above the set minima. The two-track system of
regulation reflects the American labor law’s commitment to settling distributional issues
through private bargaining and removing such issues from the political process.

The basic framework of today’s labor and employment law originated in the New Deal
period and was tailored to the job structures of that era. In that era, which I call the
“industrial era,” large firms organized their work forces into a set of practices that has
come to be termed “internal labor markets.” The term “internal labor market” is used
to distinguish these practices from the neoclassical ideal of a large impersonal external
labor market in which buyers and sellers contract freely and repeatedly for jobs of
all types.  

The internal labor market job structures of the industrial era developed in the early
and mid-twentieth century, based on the teachings of the scientific management
and personnel management schools of thought. In internal labor markets, jobs are
organized along rigidly defined lines of promotion, called job ladders. Workers are
hired at the lowest rungs and then advanced, step by step, throughout their careers.
The internal labor market job structure assumed a long-term relationship between
the employee and the firm. It also assumed that job tasks were minutely delineated
and carefully arranged so that each job provided the training for the job on the next
rung. Workers tended to stay within a particular department and on a single promotion
line, and had little lateral mobility within or between firms. They were rewarded with
longevity-based pay and benefits, and their seniority defined both their bidding rights
for higher jobs and their bumping rights in case of reductions in force.  

By the 1930s, scientific management had become business gospel, and internal labor
markets had become widespread in large industrial firms. Also in that decade, three
significant labor statutes were enacted that established a framework for governing
labor relations, which persists to this day. This framework was appropriate to the
long-term employment relationships in stable work environments that characterized
the industrial era, and it was based on the template of the mid-twentieth century
employment relationship.
Industrial era assumptions were embodied in the National Labor Relations Act. For example, a central aspect of the statutory scheme is the concept of a "bargaining unit." Under the statute, if there is a sufficient showing of interest by workers in a particular workplace, the National Labor Relations Board (the NLRB or the Board) determines the "appropriate unit" and conducts an election among employees working in the unit to determine whether a majority favor the union. If the union wins the election, the union is certified and becomes the exclusive representative of the unit for purposes of collective bargaining. Once certified, the employer and the union have a duty to bargain for a collective agreement that will govern the terms and conditions of employment for all workers in the unit, regardless of whether the employees are union members or not. Any contract concluded between the union and the employer applies to all jobs in the unit. The terms and benefits applied to the job—they do not follow the worker to other jobs when they leave the unit. At the same time, workers in the unit lose their right to take collective action apart from their certified representative, and the union has a duty to represent fairly all employees in the unit—those that support the union and those that do not.

The agency that administers the Act, the National Labor Relations Board, determines on a case-by-case basis what constitutes an appropriate bargaining unit. The Board does so by attempting to define units of employees who share a "community of interest." Some of the factors the Board uses to determine whether there is a community of interest are: similarity in kinds of work performed, similarity in compensation, types of training and skills required, integration of job functions and commonality of supervision. Under the community of interest test, bargaining units tend to have static job definitions and clear department boundaries. Yet, much of today’s work involves networks across multiple establishments or multi-employer tasks, defying traditional bargaining unit definitions. Thus the NLRB’s approach to bargaining unit determination is in tension with cross-utilization and the blurring of boundaries typical of work practices today.

The bargaining unit focus of the NLRA also means that terms and conditions negotiated by labor and management apply to the jobs in the unit rather than to the individuals who hold the jobs. As individual workers move between departments, units, or firms, their labor contracts do not follow them. Yet, today individuals experience considerable movement in their work lives, both within firms, between firms, and in and out of the labor market. As a result, in today’s world of frequent movement, union gains are increasingly ephemeral from the individual’s point of view.
Another way in which the NLRA assumes the existence of industrial era job structures is found in the scope of the Act’s coverage. The Act only provides protections for those individuals who fall within the statute’s definition of an “employee.” Individuals who work for multiple employers or the wrong kind of employer can easily fall outside the protection of the statute. Agricultural laborers, domestic workers, supervisors and independent contractors are explicitly excluded from the Act, as are government employees and employees covered by the Railway Labor Act.\(^2\) There are additional NLRB-made exclusions for managerial and confidential employees.\(^3\) Furthermore, employees who have some supervisory authority over others, or who have managerial decisions delegated to them, are excluded from coverage.\(^4\) In today’s workplace, in which hierarchies have been flattened and decision-making authority has been delegated downward, the supervisory and managerial exclusions deprive many low-level employees of the protections of the Act.\(^5\)

The exclusion for independent contractors has become particularly problematic. Because the test for independent contractor status is broad, many who are dependent on a particular employer for their livelihood are nonetheless classified as independent contractors and deprived of all labor law protections.\(^6\) Increasingly, employers attempt to reclassify employees and to vary their employment practices to transform their former “employees” into “independent contractors.”\(^7\) Many low-paid employees such as janitors, truck loaders, typists and building cleaners have been redefined as independent contractors even when they are retained by large companies to work on a regular basis.

The independent contractor exclusion also eliminates coverage for many part-time and short-term temporary workers.\(^8\) Such workers often work for more than one employer at a time, but are dependent upon and subject to the supervision of each employer for the time they are at work. Yet, when a worker has multiple employers, each employer often uses that fact to argue that the worker is actually an independent contractor rather than an employee. Courts often accept the employer’s own definition of a temporary worker’s status, thereby excluding a fast-growing portion of the workforce from unionization altogether.\(^9\)

One area in which the bargaining unit focus of the NLRA has been particularly out of step with labor market reality concerns the Act’s treatment of long-term temporary employees. Since the 1980s, temporary employment has been the fastest growing portion of the labor market. Temporary employees who work for staffing agencies are often given long-term placement at particular user firms. There, the user firm
supervises the work of the temp on a day-to-day basis, and the temp works alongside the firm’s regular employees, with the same skills, duties and job classifications. In this triangulated employment relationship, the NLRB has considered both the temporary agency and the user firm to be joint employers of the temporary employee.

In 1990, the NLRB ruled that long-term temporary employees could not be included in a bargaining unit with a user-employer’s regular employees unless both the provider-agency employer and the user-employer consented.\textsuperscript{28} Thereafter, the Board refused to consider any unit that combined temporary and regular employees, absent consent of both employers.\textsuperscript{29} Because it is highly unusual for an employer to consent to its employees forming a union, the dual consent requirement made it virtually impossible for temporary workers to unionize together with permanent workers they work alongside.\textsuperscript{30} Rather, if they want to unionize, they must do so together with the other workers employed by their temporary agency. Yet agency temporary workers are dispersed and have little contact with each other. Thus, as a practical matter, temporary workers lack representation or a collective voice under the labor law.

The U.S. labor law creates a strict divide between labor law, which pertains to collective employee rights, and employment law, which pertains to individual employee rights. Under the bifold system, core labor policy is based on bargaining and contract, with a peripheral role for legislation that establishes employment terms. Yet many of the initial employment law entitlements assume the existence of an on-going employment relationship. For example, the New Deal social security and unemployment insurance programs were not universal in their coverage. Rather, they tied crucial social insurance protections to employment, thereby reinforcing the bond between the employee and the firm. Furthermore, they did not provide mandatory and universal health insurance. Thus workers were left to obtain health insurance from individual employers, usually as a product of labor-management negotiations.

Other types of employment law protections also assume an employment relationship and hence are not available to persons designated “independent contractors.” Independent contractors are not covered by minimum wage, workers compensation, unemployment compensation, occupational safety and health laws, collective bargaining laws, social security disability, anti-discrimination laws, or any of the other employment protections discussed above.\textsuperscript{31} Unlike Europe and Canada, in the United States there have not been legislative efforts to create an intermediate category between “employee” and “independent contractor” that would give atypical workers some of the employment...
Because the labor and employment laws were tailored to the job structures of the industrial era of the twentieth century, they have become obsolete as internal labor markets have declined in importance and new ideas about how to organize work have generated new work practices throughout American enterprises. Job security in the private sector, in the form of long-term attachment between a worker and a single firm for the duration of the worker’s career, is rapidly declining. Today workers expect to change jobs frequently and employers engage in regular churning of their workplace, combining layoffs with new hiring as production demands and skill requirements shift. “Regular” full-time employment no longer carries the presumption of a long-term attachment between an employee and a single firm with orderly promotion patterns and upwardly rising wage patterns. At the same time, employers encourage employees to manage their own careers and not to expect career-long job security. In addition, there has been an explosion in the use of atypical workers such as temporary workers, on-call workers, leased workers and independent contractors.

A new employment relationship is emerging to replace the industrial era internal labor markets. Today’s world of specialty production and knowledge work has spurred the development of new job structures, the job structures of the “digital era.” In the digital era, employees do not have long-term job security with a particular employer. Some of the terms of the new employment relationship are as follows:

First, instead of job security, employers today explicitly or implicitly promise to give employees “employability security”—that is, opportunities to develop their human capital so they can prosper in the external labor market.

Second, the new employment relationship places emphasis on the worker’s intellectual and cognitive contribution to the firm. Unlike scientific management, which attempted to diminish or eliminate the role of workers’ knowledge in the production process, today’s management theories attempt to increase employee knowledge and harness their knowledge on behalf of the firm.

Third, today’s employment relationship involves compensation systems that peg salaries and wages to market rates rather than internal institutional factors. The emphasis is on offering employees differential pay to reflect differential talents and contributions.
Fourth, as part of the new employment relationship, firms now also provide employees with opportunities to interact with a firm’s customers, suppliers and even competitors. Regular employee contact with the firm’s constituents is touted as a way to get employees to be familiar with and focused on the firm’s competitive needs, and at the same to raise the employees’ social capital so that they can find jobs elsewhere.

And finally, the new relationship involves a flattening of hierarchy, the elimination of status-linked perks and the use of company-specific grievance mechanisms.

While the new employment relationship does not depend upon long-term employment, attachment or mutual loyalty between the employee and the firm, it also does not dispense with the need for engaged and committed employees. Indeed, firms today believe that they need the active engagement of their employees more than ever before. They want not merely predictable and excellent role performance, but what has been described as *spontaneous and innovative activity that goes beyond role requirements.* They want employees to commit their imagination, energies and intelligence on behalf of their firm.

Today’s valuation of employees’ cognitive contribution stands in direct contrast to the scientific management approach. Under scientific management, workers were not expected to gain or use knowledge in their jobs. Knowledge was a monopoly tightly held by management. Today, firms believe that they can acquire a competitive advantage by eliciting and harnessing the knowledge of their employees. According to Fortune magazine editor, Thomas Stewart, “Information and knowledge are the thermonuclear competitive weapons of our time.”

The emerging employment relationship has two diametrically opposed consequences. On the one hand, it creates a more interesting work environment and offers workers more autonomy and freedom than did the industrial era job structures. Yet on the other hand, for many it creates uncertainty, shifts risk and fosters vulnerability. Some of the groups that are disadvantaged in the new work regime are easily identified. For example, older workers caught in the transition are heavy losers. Having been led to expect a good job and a secure future, they instead discovered that their expectations were chimeral. Another group that has not fared well is the low-skilled—those who have neither the necessary training nor the ability to reinvent themselves, retool and adapt to new labor market demands. A third group is the risk-adverse—those who were comfortable in internal labor markets and lack the desire or initiative to seek out opportunities, to network and to build their own careers.
In addition to the older, the unskilled and the risk-adverse, all workers now face heightened risks at certain times in their working lives. Employees can expect to have episodic jobs, sometimes as regular employees, sometimes as temporary workers, and sometimes as independent contractors. Given the churning and constant change that characterizes the new workplace, all face a high likelihood that their working lives will be peppered by occasional periods of unemployment. Therefore, every worker requires a reliable safety net to ease the transitions and cushion the fall that they are likely to encounter in today's boundary-less workplace.

The foregoing historical perspective brings us to the question, what will the labor and employment law look like in the future? In the past, labor and employment laws were enacted as the result of pressure from organized labor and social reformers to ameliorate the vulnerabilities and injustices that occur in the operation of the labor market. The problem today is that the labor and employment laws no longer provide redress for the most pressing problems of workers. The changing nature of work has caused new problems to arise in the operation of the labor market, problems that call for new kinds of regulatory interventions. Today workers move frequently between firms and within firms, so bargaining-unit based unionism gives little protection. And the employment laws do not give adequate protection to the individuals who move in and out of the labor market, or who do not have a typical relationship with a single employer.

There are two possible scenarios for the future of labor law. One scenario is that labor law will continue to atrophy, unions will continue to decline and individual employee rights will be chipped away through the combined processes of narrowing judicial construction of existing rights, the development of a robust waiver doctrine whereby employees will have rights on paper but not in practice, pressures from globalization for lower labor standards and a slow erosion of specific monetary standards through inflation. This scenario is a likely one given the declining power of unions at the legislative level that results from labor's declining numerical strength. Union political power is necessary to pressure politicians to maintain employment standards at current levels or raise them higher. In this first scenario, worker rights will decline in all the respects just mentioned, and we will see a return to the laissez-faire labor regulation of the pre-Wagner Act era.

The other scenario is that labor laws will evolve in a way that represents a marked break with the present in order to address the needs and concerns of individuals in the
new workplace. I predict that changes will come in some or all of these respects:

- a partial collapse of the distinction between labor law and employment law;
- an expanded focus on the legislative front rather than on collective bargaining to set employment conditions;
- an expansion of collective bargaining to new groups, such as independent contractors, atypical workers, immigrants, unemployed workers and geographically-defined groups;
- a broadening the field of labor and employment law to include all issues of concern to working people, such as health care policy, training and education, welfare, intellectual property protection, pensions and social security, housing policy and other areas of social law; and
- the creation of a new type of social safety net to focus on the problem of transitions and gaps in people’s labor market experiences.

As stated above, the U.S. system of employment regulation has maintained a distinction between the collective bargaining rights for unionized workers and individual employment rights for other workers. Though this distinction sounds fixed in theory, there has in fact always been a permeable boundary between these bodies of regulation.44

Recent developments have further challenged this distinction. Increasingly, workers with individual employment law claims have brought their claims in a collective form, either as class actions under most employment statutes, or as “collective actions” under the Fair Labor Standards Act. Class actions have long been a feature of employment discrimination litigation, but now they have spread to other types of employment law claims. Collective actions under the FLSA are similar to class actions, but in some respects the requirements for a “collective action” are easier for plaintiffs to satisfy than those for a class action under Rule 23 of the Federal Rules of Civil Procedure.45 Collective employment litigation has been brought in both state and federal courts, alleging violations of both state and federal labor laws.46

Employment class actions occupy the vast majority of work of management-side employment law firms. As one observer writes, “A sample of 150 FLSA ‘collective action’ cases prosecuted by the Department of Labor as of January 2005 reads like a Who’s Who of corporate America, including Wal-Mart (seven times in the previous
five years); Bed, Bath & Beyond; Nortel Networks; Safeco Insurance Companies (twice); Pep Boys; Electronic Arts, Inc.; Minolta Business Solutions; Countrywide Credit Industries; Conseco Finance Corp.; NBC; Ameriquest Mortgage Co. (three times); First Union Corp.; and, Perdue Farms. Public entities being sued included the City of Louisville and the Chicago Transit Authority. The majority of these claims were misclassification cases, mostly for unpaid overtime. This list does not include private collective employment actions brought by individual or groups of employees.

Employment law collective actions can result in sizeable damage awards. For example, in recent wage and hour suits in California alone, the Coca-Cola Bottling Company settled a case for $20.2 million, Bank of America settled for $22 million and Rite Aid Corp. settled for $25 million. In 2002, United Parcel Service agreed to pay $18 million to settle a similar suit on behalf of misclassified supervisors. The same year, Starbucks Corp. also paid $18 million to settle two class action suits on behalf of current and former managers and assistant managers in California who claimed that they had been misclassified as “exempt” employees and thereby denied overtime compensation.

Collective employment litigation, whether brought as class actions or as FLSA collective actions, is an expanding form of collective action in an era of declining union activity. While such actions do not foster the experience of solidarity and collective empowerment that characterize unionization efforts and strikes, they share some features with other conventional forms of collective action. They reflect a shared sense of work-related wrong and they define a group of workers—the class—as having a shared interest. They also operate through representatives, the named plaintiffs and the class counsel, who speak to management for the workers and, at least in theory, represent their interests. Collective employment actions are greatly feared by management because, apart from their potential exposure, the suits are on-going disputes with incumbent disgruntled employees who are potentially poisonous to general workplace morale.

Class actions also have some obvious and significant differences with conventional unionization efforts. First, they do not involve the type of mobilization that typically occurs in a union drive. Furthermore, they do not aim to form lasting organizations nor do they offer the prospects of an on-going bargaining relationship between workers and an employer over the whole range of issues involved in the employment relationship. They take a long time to run their course, but essentially they are one-shot, single issue challenges to a company’s employment practices. And they also do
not generally foster the type of bonds of solidarity on which conventional unionism relies.\(^{51}\) And finally, they seek to vindicate pre-existing statutory rights, not to define the normative rules that shall govern the workplace. That is, unlike collective bargaining, they are not an exercise in labor-management self-regulation.

It is interesting to note that the features of collective employment actions that distinguish them from collective bargaining parallel the broad changes in the employment relationship described above. In collective litigation, relatively atomistic employees come together to fight on one issue. Some class members may not be employed by the defendant at the time of the lawsuit, and the members of the class often have never met each other. Once the suit is over, any bonds of solidarity dissolve. This one-shot activism is compatible with the mobile, self-contained knowledge worker that is the paradigm of today's worker. Hence it is possible that ex post single issue workplace governance is the form that collective action will continue to take in the future.\(^ {52}\)

Despite the differences between collective employment litigation and collective bargaining, as unions decline collective litigation has become an important venue for the protection of employment rights. Their profusion suggests that this may be an important form of employee collective action for the future. If that is the case, then the specific legal requirements of maintaining a collective legal action—whether a class action or a FLSA collective action—will come under increased scrutiny. For example, in some employment discrimination litigation, courts have begun to consider how to apply Rule 23(b)'s commonality and typicality requirements to workplaces in which management authority is diffuse and delegated to lower-level supervisors.\(^ {53}\) This issue is posed presently in the behemoth employment discrimination case, *Dukes v. Wal-Mart Stores*, involving 1.5 million present and former Wal-Mart employees.\(^ {54}\) There are also debates about the requirement in FLSA collective actions that class members "opt in" rather than "opt out" as is permitted under Rule 23.\(^ {55}\) As employment class actions continue to proliferate, these procedural requirements will take on added significance.

Another feature of collective employment litigation is the involvement of unions. More and more, unions are financing and otherwise assisting unorganized workers in mounting employment class actions. For example, the United Food and Commercial Workers Union has been actively involved in wage and hour suits against Albertson's grocery chain, Tyson Foods, Perdue Farms and the Nordstrom retail chain.\(^ {56}\) The Writers' Guild sponsored several wage and hour class action lawsuits against television reality shows even though the employees involved were not represented by the union.\(^ {57}\)
Some have argued that by assisting these types of actions, unions can gain a foothold in unorganized workplaces that could lead to greater organizing success down the road. While there is no evidence to date that this has occurred, it remains a hopeful prospect for a labor movement that is experiencing hemorrhaging losses.

Before we can conclude that collective employment actions are either a substitute for actual unionization or a foot-in-the-door method to revitalize the union movement, it is necessary to look at some legal issues that are waiting in the wings. One issue that has arisen is whether a union, by giving unorganized workers financial assistance in the form of legal representation in employment litigation, is giving an unlawful benefit to improperly influence workers’ choice whether or not to unionize. Some court decisions have held that when a union finances an employment litigation, it is an unlawful payment of benefits and hence grounds to set aside a union election.58

Another issue that might arise is whether a union that participates in the negotiation of a settlement of an employment class comprised of unorganized workers is acting in a representative capacity without having attained majority status. In such a case, its actions would also violate the statute. If union involvement in employment class actions is to be an important tactic in the future of the union movement, the labor law will need to address these issues.

In recent years, as unions have declined, more statutory employment rights have been created that are applicable to all workers. In addition, the nature of legislated individual employment rights has shifted from a floor to a baseline. That is, the more recent employment standards are not designed to set bare minima, but to set an adequate baseline level of protection. For example, the Occupational Safety and Health Act59 imposes a general duty on the employer to provide each worker a work environment that is free from identified hazards.60 Similarly, workplace privacy protections and employment discrimination legislation is designed to ensure individuals a workplace that is free of discrimination and respectful of employee privacy. This is not to say that these and other employment rights are set at an optimal or even a truly adequate level. For example, the Family and Medical Leave Act mandates a minimal period of leave for child-bearing, but does not mandate pay replacement for the period of the leave. But unlike the original New Deal employment rights, the more recent statutory rights are intended to apply to a majority of workers, not merely those at the margins of subsistence.

The change in the nature of employment rights and the increase in rights for all
employees represent a shift in the locus of employment regulation away from collective bargaining and toward the state. The shift from collective bargaining to legislation does not necessarily signal the end of unionism, but rather could auger a change in union strategy and tactics. Unions may shift their focus from exerting employer-specific pressure to exerting pressure in the political arena, including federal, state and local. This would represent a significant departure from the U.S. labor movement’s traditional position, dating back to Sam Gompers in the 1890s, that union pressure was most effective in the economic realm rather than in the political realm. But today, as discussed above, employer-centered union pressures are rendered less effective than they were in the past because employees have little attachment to either a specific employer or a particular craft group. Should unions concentrate more on the political realm, they will also be forced to articulate their goals in the vernacular of the public interest rather than as the demands of a special interest group. Hence, they may reclaim their role as spokesperson for the working population generally.

In terms of the future of employment law, we can expect not only more employment laws, but also more controversy about them. Campaign finance rules regarding union campaign contributions, and labor law rules regarding use of union funds for political purposes, have generated enormous amounts of litigation in recent years, and we can expect those controversies to intensify.61

There is evidence that employees feel they need unions, but not necessarily the unions that now exist.62 Given the decline of worker-firm attachment, workers need organizations that further their joint interests but that are not pegged to a particular employer. Because workers move frequently within and between firms throughout their working lives, there needs to be a mechanism for workers to deploy their collective power to negotiate conditions across employers.

At the present time, there are some new types of organizations that attempt to engage in bargaining with multiple employers in different industries and utilizing workers with differing skills. For example, in many cities, unions have worked with community groups to enact living wage ordinances to improve labor standards for low-wage public sector employees.63 Presently, there are city-wide living wage ordinances in Baltimore, Los Angeles and other places as a result of area-wide political pressures by community and labor groups. Although such ordinances are limited to public sector employees, they suggest a new form of bargaining for workers across industries on a locality-wide basis. We could envision city ordinances that set industrial safety codes, mandate paid

Expanding Collective Bargaining to New Groups
family leave, require employers to provide health insurance and address other issues that are part of the shared needs of all working people in the area.

In a similar vein, in Los Angeles, San Antonio and some other cities, unions and community groups have worked together to negotiate agreements with city authorities and private investors to provide job creation, job training, affordable housing, social services, public parks and other community improvements in exchange for support for development projects. There have also been multiple-employer organizing efforts of immigrant workers within particular sectors. In many cities, worker centers have developed to inform low wage workers, often immigrants, of their legal rights.

The present labor law does not easily accommodate area-wide multi-employer, multi-sector bargaining, particularly when it involves union-community partnerships on one side, and multiple employers and city agencies on the other. However, organizations that engage in such efforts could provide important benefits for workers in today's labor market. Although workers change jobs more than in the past, they usually find new jobs in the same geographic area of their previous jobs. Hence, it would be desirable for the labor law to facilitate area-wide bargaining on such issues as minimum pay levels, health and pension benefits, leave policies, safety standards, job training programs, job transfer rights and employment benefits at the local and/or regional level. To do so, the labor law would have to abandon the present notion of bargaining units, and devise another mechanism for determining legally sanctioned bargaining rights. Proposals for geographic unions, such as put forward by Charles Heckscher, Raymond Miles and this author, can serve as a starting point.

The field of labor and employment has, until now, been seen as narrowly related to issues that arise in the employer-employee relationship in the workplace. However, given today's fluid and boundary-less workplace, issues concerning work do not always involve workers' relationships to their immediate employer. Rather, many other areas of law affect one's relationship to the labor market. These include issues such as health insurance, training and education, welfare assistance, pensions and social security. Also, there are new issues that have arisen for workers as a result of the new employment practices.

For example, one legal issue that was invisible in the past but has become prominent today is the issue of who owns an employee's human capital. Because the new employment relationship relies on employees' intellectual, imaginative and cognitive
contribution to the firm, employers put a premium on human capital development and knowledge-sharing within the firm. Yet the frequent lateral movement between firms that typifies the new relationship means that when an employee leaves one employer and goes to work for a competitor, there is a danger that proprietary knowledge will go too. Increasingly, the original employer, fearing that valuable knowledge possessed by the employee will fall into the hands of a competitor, will seek to prevent the employee from taking the job or utilizing the valuable knowledge. Yet employees understand that their employability depends upon their knowledge and skills, so they assume that they can take their human capital with them as they move between jobs. As a result of these conflicting perspectives, legal disputes about post-employment covenants and trade secrets have increased exponentially, making them the most frequently litigated issues in employment law.

In the future, it will be important to create a new type of social safety net, one tailored to the vulnerabilities of today’s workplace. Because most workers will experience discontinuities in their labor market experiences, they need provision for gaps and transitions. They need portable health benefits, lifetime training and retraining opportunities, universal and adequate old age assistance, and other forms of assistance for individuals who are in periods of transition between jobs or changing careers.

To date, neither our welfare laws nor our labor and employment laws have focused on the problem of transition assistance. However, the issue has been actively considered in Europe. In 1999, the European Commission convened a distinguished group of labor relations experts to consider the implications of the changing nature of work on labor regulation in Europe and to devise proposals for reform. In 2000, the group, chaired by Professor Alain Supiot, issued a report that described a changing employment landscape as a result of the movement away from internal labor markets toward more flexible industrial relations practices. The Supiot Report called for new mechanisms to provide workers with “active security,” by which they mean mechanisms that equip individuals to move from one job to another. It contained a number of suggestions for changes in the institutions regulating work to provide active security. Its most visionary proposal was for the creation of “social drawing rights” to facilitate worker mobility and to enable workers to weather transitions. Under the proposal, an individual would accumulate social drawing rights on the basis of time spent at work that could be used for paid leave for purposes of obtaining training, working in the family sphere, or performing charitable or public service work. It would be a right that the individual could invoke on an optional basis to navigate career transitions, thereby giving
flexibility and security in an era of uncertainty. As Supiot writes, “They are drawing rights as they can be brought into effect on two conditions: establishment of sufficient reserve and the decision of the holder to make use of that reserve. They are social drawing rights as they are social both in the way they are established…and in their aims (social usefulness).”

The report makes an analogy to sabbatical leaves, maternity leaves, time off for union representatives and training vouchers to observe that “we are surely witnessing here the emergence of a new type of social right, related to work in general.” Social drawing rights, it is said, would smooth career transitions and give individuals the resources to retool and to weather the unpredictable cycles of today’s workplace.

In the United States, we have precedents for the concept of paid time off with re-employment rights to facilitate career transitions or life emergencies. We have long permitted paid leaves for military service, jury duty, union business and other socially valuable activities. Some occupations also offer periodic sabbatical leaves. The concept is also built into the idea of temporary disability in state workers compensation and other insurance programs, which provide compensation and guarantee re-employment after for temporary absences. The recent Parental Leave Act extends the concept of leave time to parenting obligations. These programs all reflect and acknowledge the importance of subsidized time away from the workplace to facilitate a greater contribution to the workplace. They could serve as the basis for developing a more generalized concept of career transition leave, or to use more familiar parlance, a workplace sabbatical.

A workplace sabbatical would be a right, accrued by time spent in the labor force, to paid leave for the purpose of retooling, retraining and repositioning oneself in the labor market. This right should be made a part of the contract of employment similar to a right to unemployment compensation. The workplace sabbatical right should not be an implied-in-fact term of the contract of employment—that is, it should not depend upon an employer implicitly promising employability, training and networking opportunities, and it should not be waivable. Rather, the right to a workplace sabbatical should be an implied-in-law term that it grows out of the recognition that workers today are vulnerable to changing technological demands, and need opportunities to change and develop their human capital as they face a lifetime of job transitions. The justification for imposing such a term is that it tracks the normative as well as practical reality of today’s workplace.
The workplace is changing and the labor and employment laws will change as well. Workers today are forced to bear many new risks in the labor market—risks of job loss, wage variability, benefit gaps, skill obsolescence and intermittent prolonged periods of unemployment. Currently our labor and employment laws do not address these problems, either for regular or for atypical workers. The changing nature or work has rendered much of the labor and employment law framework obsolete, and a new framework will be created to take its place. It remains to be seen whether the new framework will be a free market framework of laissez-faire capitalism, or whether it will be the creation of a new type of rights and safety net that enables workers to thrive in the new workplace.
Professor of Law, UCLA School of Law. This piece is excerpted from Katherine V.W. Stone, The Future of Labor and Employment Law in the United States, in ENCYCLOPEDIA OF LABOR AND EMPLOYMENT LAW AND ECONOMICS (Kenneth G. Dau-Schmidt et al. eds., forthcoming 2009).


7 PETER DOERINGER & MICHAEL J. PIORE, INTERNAL LABOR MARKETS AND MANPOWER ANALYSIS (1971).

8 HOW LABOR MARKETS WORK: REFLECTIONS ON THEORY AND PRACTICE BY JOHN DUNLOP, CLARK KERR, RICHARD LESTER, AND LLOYD REYNOLDS (Bruce E. Kaufman ed., 1988).


10 DOERINGER & PIORE, supra note 7, at 1–3; STONE, WIDGETS TO DIGITS, supra note 6, at 53–56; PAUL OSTERMAN, INTERNAL LABOR MARKETS 2 (1984).


12 For a detailed description of the 20th century employment relationship, see STONE, FROM WIDGETS TO DIGITS, supra note 6, at 27–63.


14 Under the National Labor Relations Act (NLRA), a union can also be designated as an exclusive representative by means of an employer grant of recognition after a showing of a card majority or other convincing evidence of majority support. See NLRB v. Gissel Packing Co., 395 U.S. 575, 592 (1969). But certification as a result of a Board-sponsored election is the preferred method of obtaining representative status under the NLRA. See id. at 596.


17 See Vaca v. Sipes, 386 U.S. 171, 177, 182 (1967); Steele, 323 U.S. at 200–02.


22 See NLRB v. Health Care & Ret. Corp., 511 U.S. 571, 578 (1994) (finding that charge nurses are “supervisors” under the statute because they assign work to nurse’s aides); NLRB v. Yeshiva Univ., 444 U.S. 672, 679–82 (1980) (holding that university professors are “managers” for purposes of exclusion because they exert collective decision-making authority in hiring, curriculum, and other matters).

23 See, e.g., NLRB v. Ky. River Cmty. Care, Inc., 532 U.S. 706 (2001) (finding certain nurses to be “supervisors” even though they had no subordinates and had no authority to hire, fire, promote, reward, or evaluate other employees).

24 See generally Katherine V.W. Stone, Legal Protections for Atypical Employees: Employment Law for Workers Without Workplaces and Employees Without Employers, 27 BERKELEY J.


26 In the 1947 amendment to section 2(2) of the NLRA, 29 U.S.C. § 152(2), Congress rejected an “economic reality” test in favor of a common law test for determining independent contractor status. However, the Board and courts of appeal have often differed as to what that test requires.

27 See, e.g., Clark v. E.I. DuPont de Nemours & Co., 105 F.3d 646 (4th Cir. 1997) (per curiam); Abraham v. Exxon Corp., 85 F.3d 1126, 1132 (5th Cir. 1996). But see Vizcaino v. United States Dist. Court, 173 F.3d 713, 724–25 (9th Cir. 1999) (rejecting an employer’s assertion that employees are independent contractors for purposes of eligibility for a stock purchase plan).


30 In 2000, the NLRB reversed its former position and held that regular employees and temporary employees could be in the same bargaining unit so long as they shared a community of interest. In re M.B. Sturgis, Inc., 331 NLRB 1298 (2000). However, in 2004 the NLRB again reversed itself in the case of Oakwood Care Center and N & W Agency, and reinstated the dual consent requirement for temporary worker organizing efforts. 343 NLRB No. 76 (2004).


33 According to the United States Department of Labor’s Current Population Survey, job tenure for men between 55 and 65, measured as the average time with a given employer, declined from 15.3 to 10.2 years between 1983 and 2002. For men between 45 and

34 For a detailed description of the changing workplace, see STONE, FROM WIDGETS TO DIGITS, supra note 6, at 87–114.


37 ROSABETH MOSS KANTER, ON THE FRONTIERS OF MANAGEMENT 175 (1997) (reporting that the tide is moving “toward more varied individual compensation based on people’s own efforts”).

38 For example, one of the most touted practices of Total Quality Management is that “management should seek to create conditions whereby every worker, at least from time to time, sees and talks with real customers, with actual users of the company’s product or service.” ERIC E. ANSCHUTZ, TQM AMERICA: HOW AMERICA’S MOST SUCCESSFUL COMPANIES PROFIT FROM TOTAL QUALITY MANAGEMENT 53 (1995).


42 THOMAS A. STEWART, INTELLECTUAL CAPITAL, supra note 36.

43 For example, a case study of white collar workers laid off at IBM and Link Aerospace in Binghamton, New York—two companies known for their paternalistic long-term employment relationships—concluded that “downsizing and displacement change the expectations about the relationships among workers and between employers and workers.” Charles Koeber, Corporate Restructuring, Downsizing, and the Middle Class: The Process and Meaning of Worker Displacement in the “New” Economy, 25 QUALITATIVE SOC. 217, 219 (2002).

44 The labor law Section 301 preemption doctrine serves as the primary traffic cop that directs individuals with work-related disputes to one body of law or the other. However, the
preemption doctrine itself has been an evolving and changing set of rules, so that some individual rights can be vindicated by individuals who have union contracts, and some cannot. See Katherine V.W. Stone, The Legacy of Industrial Pluralism: The Tension Between Individual Employment Rights and the New Deal Collective Bargaining System, 59 U. CHI. L. REV. 575 (1992).

45 Collective actions under the FLSA are provided for at 29 U.S.C. § 216. They do not have the same stringent requirements for numerosity and typicality that are imposed by Rule 23 of the Federal Rules of Civil Procedure for class actions, so it is easier for a collective action to be maintained under the FLSA.


50 This point was made to me in conversation by the head of employment litigation for one office of the Jackson Lewis law firm, one of the largest employer-side employment law firms in the United States.


52 I am grateful to Fred Tung for this insight.


54 Dukes v. Wal-Mart Stores, Inc., 222 F.R.D. 137 (N.D. Cal. 2004), aff’d, Dukes v. Wal-Mart...
Stores, Inc., 474 F.3d 1214 (9th Cir. 2007), withdrawn and superseded by Dukes v. Wal-Mart Stores, Inc., 509 F.3d 1168 (9th Cir. 2007) (affirming lower court’s class certification).


56 Hechler, supra note 48.


58 See Freund Baking Co. v. NLRB, 165 F.3d 928 (D.C. Cir. 1999); Nestle Ice Cream Co. v. NLRB, 46 F.3d 578 (6th Cir. 1995); see also United Food and Commercial Workers Union, Local 120 (Wal-Mart Stores), Case No. 32-CB-5757-1 (2004), 2004 WL 2414080. But see 52nd St. Hotel Assoc., 321 NLRB 624 (1996) (refusing to find union’s conduct in bringing FLSA lawsuit on behalf of unorganized employees to be objectionable or grounds to set aside election). See generally Catherine L. Fisk, Union Lawyers and Employment Law, 23 BERKELEY J. EMP. & LAB. L. 57 (2002) (discussing legality of union representation of non-union workers in wage and hour litigation).


60 Id. § 654(a)(1).


67 See Raymond E. Miles, Adapting to Technology and Competition: A New Industrial Relations System for the 21st Century, 31 CAL. MGMT. REV. 9, 23–25 (1989) (advocating a turn to geographically based unionism); Charles C. Heckscher, Associational Unionism, in THE NEW UNIONISM: EMPLOYEE INVOLVEMENT IN THE CHANGING CORPORATION 177 (1988); STONE, FROM WIDGETS TO DIGITS, supra note 6, at 217–39. For details about the labor law changes that such proposals would require, see id. at 237–39.


69 SUPIOT ET AL., BEYOND EMPLOYMENT, supra note 66, at 56.

70 Id.
Steven Bank teaches Introduction to Federal Income Taxation, Taxation of Business Enterprises, Corporate Tax Policy, the Tax Policy and Public Finance Colloquium, Tax Aspects of Mergers and Acquisitions, and a seminar on Corporate Tax Policy. His research generally explores the taxation of business entities through the lens of legal and business history. Professor Bank has published numerous articles and chapters in the fields of business taxation, tax policy, and tax history and is the co-author or editor of two forthcoming books on the taxation of business enterprises.
Kirk Stark teaches Federal Income Taxation, Taxation & Distributive Justice, Financing State and Local Government, and the first-year Property course. In addition, he has regularly served as faculty coordinator of the UCLA Colloquium on Tax Policy & Public Finance. Professor Stark was elected Professor of the Year by the law school graduating classes of 1999 and 2002. In 2003, he received the University Distinguished Teaching Award. Professor Stark writes in the areas of tax policy and public finance. Much of his recent scholarship has examined the American system of “fiscal federalism” and considers how best to allocate fiscal responsibilities among federal, state and local governments. Stark was recently chosen as a 2006-2007 Fellow for the UCLA Center for American Politics and Public Policy.
WAR AND TAXES†

Steven A. Bank & Kirk J. Stark

In the early summer of 1967, veteran Washington journalist Peter Lisagor met with a senior Republican senator to discuss the deteriorating situation in Vietnam. The war had divided the country, triggering massive antiwar demonstrations in several major cities, and the senator agreed to talk only on condition of anonymity. But the topic of discussion was not troop levels or moral arguments over the U.S. presence in Indochina. Rather, the senator wanted to talk about something more mundane: taxes. As Lisagor later explained in a Los Angeles Times article, "Absence of Sacrifice at Home Spurs Guilt Feeling over War," the GOP senator considered taxes a question of conscience. "I went to the beach with my son and his children a few weeks ago," the senator explained, "and there we were, enjoying ourselves as if we didn't have a care in the world. We had no sense of a war, no sense of sacrifice. Yet this war is already bigger than Korea. I'll go for a tax increase now."

A generation later, the senator's question of conscience has resurfaced in public debate. On March 19, 2003, the Bush administration launched Operation Iraqi Freedom, a military campaign to overthrow dictator Saddam Hussein. Administration officials defended the action as part of a broader "war on terror," including Operation Enduring Freedom in Afghanistan, which began shortly after the al Qaeda attacks of September 11, 2001. From that point forward, the United States has been actively waging a costly overseas military operation. Within six years, the Department of Defense had confirmed a total of 4,018 U.S. fatalities in Iraq and Afghanistan. And according to estimates from the Congressional Budget Office, by 2007, the budgetary cost of operations in the two countries exceeded $500 billion.

Yet despite the country's great loss of blood and treasure, there is little sense of sacrifice on the homefront. Indeed, in its first six years, the Bush administration has requested, and Congress has approved, a series of major tax cuts. Lawmakers have lowered and flattened rates for the individual income tax, initiated a repeal of the estate tax, eased the burden on capital gains and corporate dividends, reduced the so-called marriage penalty, and enacted a slew of new deductions, credits and other special-interest provisions. When combined with a steady increase in military, domestic and entitlement spending, these cuts have turned a projected $5.6 trillion surplus over the
This contrast—between an active war effort on one hand and substantial tax cuts on the other—has no precedent in American history. Beginning with the War of 1812, special taxes have supported every major military conflict in our nation’s history. Moreover, many levies have outlasted the wars they financed. Politicians like to talk about their plans for revamping the country’s tax system, but important tax reform usually happens when it must, not when it should. War has been the most important catalyst for long-term, structural change in the nation’s fiscal system. Indeed, the history of America’s tax system can be written largely as a history of America’s wars.

Enactment of the Bush tax cuts has called into question the once-axiomatic relationship between war and taxes. The historical incongruity of Congress reducing taxes while increasing spending on the war in Iraq has provided fodder to administration critics who, like the anonymous senator calling for increased taxes to pay for the war in Vietnam, have wondered publicly if the country has betrayed its tradition of wartime fiscal sacrifice. As one pundit declared in a typical statement, “in his determination to cut taxes even while waging war in Iraq, President Bush is bucking history.” Yet another bemoaned, “since 9/11, our government has asked no sacrifice of civilians other than longer waits at airplane security. We’ve even been rewarded with a prize that past generations would have found as jaw-dropping as space travel: a wartime dividend in the form of tax cuts.”

Underlying these comments is an inescapable fact: the United States has a strong tradition of wartime fiscal sacrifice, and the Bush tax cuts mark an abrupt departure from that tradition. As we hope to illustrate, however, America’s history of wartime taxation is not quite the heroic tale that many Bush critics seem to imply. Although taxes have typically gone up during times of war, the claim that “we have always accepted heavier burdens as the price those at home pay to support those under fire on the front” misses much of the complexity of American history. Indeed, as a nation, our commitment to wartime fiscal sacrifice has always been uneasy—and more than a little ambiguous. In some wars, political leaders have asked Americans to accept new taxes as the price of freedom and security. But in others, they have tried to delay, deny, and obscure the trade-off between guns and butter. And even when Americans have embraced the call for sacrifice, their elected representatives have often made room for self-indulgence, easing burdens for some constituents while raising them for others.

Exaggerating the American tradition of wartime fiscal sacrifice is understandable but...
unfortunate. History is most usable, at least for politicians, when it can be recast as a morality play. But it is most valuable, at least for the rest of us, when it honestly probes the inconvenient truths of human nature and political struggle. In our search for the historical context of current debates, we should be careful not to compare today's policies to some cardboard cutout version of an imagined past.

* * *

As political scientist David Mayhew recently observed, since its founding in 1789, the United States “has conducted hot wars for some 38 years, occupied the South militarily for a decade, waged the Cold War for several decades, and staged countless smaller actions against Indian tribes or foreign powers.” The cost of these activities has been immense, with important and lasting consequences for the tax system, the economy and the nation's political structure. By focusing on tax legislation, we hope to identify some of these consequences. But we are not interested in simply recounting statutory details. Rather, we hope to illuminate the politics of war taxation, with a special focus on the influence of arguments concerning “shared sacrifice” in shaping wartime tax policy. Moreover, we aim to shed light on a less examined aspect of this history by offering a detailed account of wartime opposition to increased taxes.

Historically, two features of wartime politics have prompted tax reform. The first is sheer necessity. There is simply no other government activity that requires as much revenue as fighting a war. Success on the battlefield requires economic resources, and taxation is the best means of marshalling those resources. While explicit taxes are not the only means of extracting resources from a nation and its people, practical limits on nontax forms of war financing (e.g., borrowing, seigniorage, conscription, expropriation) generally push tax changes onto the legislative agenda. Second, wars often create a new political atmosphere—one characterized by feelings of solidarity and shared sacrifice. Wars may foster a feeling of “civic engagement” or a “public mood” as citizens “rally 'round the flag.” Whatever term is used, war creates new political opportunities when it comes to tax policy. Taxes are never popular, but they are never more popular than during wars. In combination, these two features of wartime politics—fiscal necessity and political opportunity—set the stage for sweeping and durable tax reform.

The most compelling example of wartime fiscal sacrifice comes from World War II. In the months following the Japanese attack on Pearl Harbor, fiscal necessity and political opportunity converged to produce dramatic changes in the nation's tax
system. Though authorized by the Sixteenth Amendment in 1913 and established by statute shortly thereafter, the income tax has its modern roots in the Revenue Act of 1942. That legislation, enacted less than a year after the official U.S. entry into the war, subjected millions of new taxpayers to the income tax, converting what had long been a “class tax” to a full-fledged “mass tax.” More than just raising revenue for the war, the Revenue Act of 1942 gave rise to a whole new taxpaying culture. The federal government launched an all-out campaign to market the new tax changes, including Disney-produced animated shorts featuring Donald Duck touting the importance of “taxes to beat the Axis!” The campaign was a success. Asked in February 1944 whether they considered the amount of income tax they paid to be “fair,” a stunning 90 percent of Americans answered yes.

The experience of World War II, so important to the image Americans have of themselves and their place in the world, has no doubt also shaped our intuitions about the American tradition of wartime fiscal sacrifice. Yet in many ways, World War II is an outlier on the continuum of war tax politics. Taking a wider historical view, beginning with the nation’s founding and continuing through the present day, we observe greater heterogeneity in the country’s willingness to accept heavier burdens of taxation during times of war. While the World War II example has parallels in certain other conflicts—most notably World War I and the Korean War—the country’s political instincts have often pushed in the opposite direction, prompting Americans and their elected leaders to resist the burdens of heavy wartime taxation.

Indeed, resistance and reluctance are recurring themes in the history of American wartime taxation. In the War of 1812, for example, congressional Republicans repeatedly balked at imposing new taxes to fund “Mr. Madison's War,” with nearly disastrous consequences for the nation’s fiscal health. Their reluctance stemmed from a widespread conviction that the war would be quick and relatively painless. It also reflected no small amount of fear that new taxes might be politically disastrous for anyone who supported them. Either way, at this early stage in U.S. history, the evidence hardly supports our cherished image of selfless Americans rushing to shoulder their wartime fiscal burdens.

In the Civil War, politicians again resisted the need for fiscal sacrifice—at least initially. Eager to minimize internal opposition to the war, leaders of both the Union and the Confederacy predicted a short—and relatively cheap—conflict. Eschewing heavy taxes, they relied on other, less onerous forms of war finance, including loans. But as evidence of tangible sacrifice grew—through the loss of life, liberty, and property—that strategy
falter. The demand for fiscal sacrifice grew ever stronger, with lawmakers seeking
to finance the war with taxes that spread the burden equitably among the populace.
Notably, this call for shared sacrifice accompanied the creation of a military draft, with
political leaders linking the conscription of able-bodied men with the conscription of
national wealth.

The war in Vietnam reveals a similar experience. As with the War of 1812 and the Civil
War, political leaders initially hoped to avoid new war taxes. The immediate political
calculus was, of course, different; Lyndon Johnson refused to ask Congress for higher
taxes to fight the war because he feared doing so might endanger his cherished “Great
Society” programs, especially among conservative Democrats who controlled the two
congressional tax-writing committees. When he eventually did submit a surtax proposal,
it was held up for almost a year because Johnson refused to agree to congressional
demands for corresponding cuts in domestic spending. Again, the historical experience
departs significantly from the popular notion of a country eager to put its fiscal muscle
behind its military might.

By highlighting this alternative tradition of wartime finance—a tradition marked by
reluctance and resistance, as well as willing sacrifice—we do not mean to minimize
the burdens that previous generations agreed to bear. The United States does, indeed,
have a tradition of wartime fiscal sacrifice. But this tradition has been more complex—
and more hotly contested—than might seem convenient for modern critics of the war
in Iraq. America’s wartime leaders, and its presidents in particular, have often been
reluctant to demand much fiscal sacrifice from their fellow citizens, at least initially.
Unwilling to risk domestic achievements, or fearful of eroding support for an unpopular
war, they have shrunk from the tough decisions that wars invariably demand. Eventually,
however, they all accepted the hard realities. Whether ardent tribunes of fiscal sacrifice
(like Franklin Roosevelt) or reluctant champions of fiscal responsibility (like Lyndon
Johnson), they all accepted the need for some sort of homefront sacrifice, as both an
economic and moral necessity.

* * *

As we complete this manuscript in early 2008, we cannot ignore its most obvious
contemporary context: is the war in Iraq somehow different from all the wars—and war
taxes—that preceded it? Despite the huge expense and the lingering nature of the
conflict, Congress and the president have refused to ask the American public for fiscal
sacrifice in the form of higher wartime taxes. Indeed, they have reduced the overall
tax burden multiple times. What accounts for this divergence from the usual practice?
In our view, three key features of the modern policymaking environment differentiate it from previous conflicts, making wartime tax cuts possible for the first time in American history.

First, the chief economic justification for wartime tax increases—fear of ruinous inflation—has been an insignificant factor during the war in Iraq. Without that economic imperative, policymakers have been free to consider unconventional wartime fiscal policies, including tax cuts. Second, significant political changes, including most notably the increased polarization of partisan elites, have resulted in the marginalization of deficit concerns and the corresponding decline in influence of so-called “deficit hawks.” As a result, the political constituency for pay-as-you-go war financing has been weaker in recent years than during any other military conflict in the nation’s history. Finally, the elimination of the military draft in 1972 removed one of the most compelling moral arguments for wartime taxes.

In every major military conflict in U.S. history, policymakers have faced the war financing decision with the prospect of disastrous inflation ever present in their deliberations. No major U.S. war has been exempt from these pressures. Reliance on currency finance during the Revolutionary War led to a collapse in the new continental currency; the War of 1812 forced commodity prices sharply upward; both the Confederacy and the Union faced pressure to increase taxes to stave off inflation; and during each of the major conflicts of the 20th century, political concern over uncontrollable price increases prompted policymakers to turn to current taxation to fund a substantial share of war expenditures. As economic historian Claudia Goldin has observed, “Every major war fought by the United States has been associated with price inflation. In fact, there are no extreme price peaks [between the years 1775 and 1975] that are not accompanied or preceded by a war.”

Given the historical record, one might even go so far as to suggest that preventing inflation has been the core concern of wartime tax policy in U.S. history. Over the past quarter century, however, the threat of inflation—and its corresponding influence on tax policy—has substantially abated. There are many reasons for today’s relatively benign inflation environment, including the downward pressure on prices exerted by the increased globalization of the economy. In addition, many attribute the low inflation rates of the past quarter century to the introduction of significant changes in the country’s monetary policy ushered in by Chairman of the Federal Reserve Board, Paul

Inflation and the Economic Imperative of War Taxes

[74]
Volcker, in the early 1980s.\textsuperscript{15}

Whatever the reason, the political consequences of a low inflation environment are unmistakable. In the past, lawmakers who opposed wartime tax increases ran the risk of being blamed for inflation and the havoc it wreaked on the economy. Today, however, there seems to be little fear among those crafting fiscal policy that their choices might endanger price stability. In one sense, therefore, there is a very simple answer to the question of why policymakers have not raised taxes to fund the war in Iraq—because, as yet, they have not been forced to do so. Indeed, having been freed from the economic imperative of avoiding inflation, policymakers have been able to reduce taxes in the face of rising war expenditures.

A second differentiating feature of the current policymaking environment is the political marginalization of concerns about federal budget deficits in recent years. In the country's previous conflicts, there has always been a strong constituency in favor of fiscal discipline and against excessive reliance on deficit financing. Concern for budget deficits reached its peak during the Korean War, when lawmakers from both parties, having experienced high inflation during World War II, were keen to avoid what they viewed as the fiscal mistakes of the past. Recall that for the fiscal year 1951 the federal government actually recorded a budget surplus, in large measure because of the tax increases enacted via the Revenue Act of 1950 and the Excess Profits Tax Act of 1950. In today's vernacular, President Truman would be considered the ultimate “deficit hawk.” In the history of American war finance, that “deficit hawk” perspective has always been given voice.

This is not to suggest that concern over budget deficits always prevailed in the formulation of tax policy during all of the country's major conflicts. Indeed, more often than not the country relied heavily on deficit financing during wartime. During World War II, for example, deficits reached as high as 30 percent of GDP, a level unlikely to ever be seen again. Yet even in World War II, policymakers took extraordinary measures to reduce the government's reliance on deficit financing. By contrast, recent tax policy has been marked by a specific rejection of deficit concerns, even as the country prepared to go to war. As Vice President Dick Cheney famously quipped in late 2002, in response to Treasury Secretary Paul O'Neill's expression of concern about the country's fiscal soundness, “Deficits don't matter. We won the midterms. This is our due.”\textsuperscript{16}
Why have deficit concerns played such a marginal role in the formulation of tax policy during the Bush years? What changes in American society account for the apparent decline in the influence of “deficit hawks,” who might have pushed tax policy in the traditional direction of tax increases during war? On this point, we believe some attention should be given to the very substantial political changes that the country has undergone since the mid-1970s. As political scientists have observed, the country’s political establishment has grown more polarized in the past three decades, with liberals becoming more liberal and conservatives becoming more conservative. This is not a loose “gestalt” type judgment made by pundits, but rather an empirical observation based on lawmakers’ roll call votes in Congress. Recent research undertaken by political scientists has shown that the policy positions of the average Democrat and the average Republican have become more widely separated since the mid-1970s. The result, as one recent study put it, is that “the moderates are vanishing from Congress.”

The consequences of a more polarized political environment for war financing decisions should not be underestimated. Because deficit hawks come disproportionately from the moderate ranks in both parties, their influence has suffered a decline that roughly corresponds with the rise of partisan polarization. Indeed, the story of the Bush-era wartime tax cuts is perhaps best understood as the triumph within the GOP of conservative “growth hawks” over the more moderate “deficit hawks.” The effect has been more pronounced in the House of Representatives than in the Senate. Recall that in connection with JGTRRA 2003, moderate Republicans in the Senate, including most notably Senators John McCain, Olympia Snowe, George Voinovich, and Lincoln Chafee, were able to hold down the overall cost of the administration’s second tax cut to $350 billion. It is noteworthy that there was no similar movement in the House, which because of redistricting is more susceptible to the polarizing trend.

In combination with the economic factor of historically low inflation rates, the political developments of increased partisan polarization and the corresponding marginalization of deficit concerns produced something of a “perfect storm” of conditions for wartime tax cuts. Any analysis of U.S. wartime tax policy would be incomplete, however, without reference to the chief moral argument for wartime tax increases—i.e., the U.S. taxpayers should share in the sacrifice borne by American soldiers on the field of battle.

As we have emphasized at various points in our analysis, a major difference between the war in Iraq and previous conflicts is the absence of mandatory
military service and the corresponding effect on the politics of wartime tax policy. Shared sacrifice has been a major theme in the politics of wartime taxes throughout the country's history. However one feels about the costs and benefits of conscription, the drafting of ordinary citizens into military service has profoundly influenced the way the country talks about the costs of war.

Conscription adds an unmistakable moral force to the arguments of those who advocate wartime tax increases and obliges opponents of higher taxes to reframe, or perhaps even abandon, their arguments. Recall how Representative Edward Little of Kansas framed his argument at the outset of American involvement in World War I: “You promised when you conscripted the youth of this country that you would conscript the wealth as well. …Let their dollars die for this country too.”20 Truman's Treasury chief, James Snyder, issued a similar admonition to the Senate Finance Committee during the Korean War, alluding once again to the “conscription” of wealth as well as men: “You passed a bill up here to draft boys of 18, to send them to war. I think it is just as important we draft some of the profits to help pay for the expenditures.” Opponents of higher wartime tax burdens have likewise reformulated their arguments to appear more sensitive to the burdens upon those drafted into military service. For example, consider Senator Russell Long's awkward argument that a “tax increase of ten times the size recommended by the president would still not begin to [equal] the sacrifice of our courageous young men fighting and dying in the swamps and jungles of Vietnam.”21

Given the frequent invocation of conscription as a justification for wartime tax increases, it seems reasonable to conclude that Americans are more willing to accept higher taxes when those burdens are framed in the context of the sacrifices of American soldiers. If so, it would appear that the elimination of the draft in 1972 and the introduction of the All-Volunteer Force shortly thereafter worked an unexpected transformation on the politics of wartime taxation. Whereas conscription made wartime taxes more likely, or at least provided an obvious and compelling argument in their favor, the introduction of a professional volunteer military force eclipsed those arguments completely. From 1973 onward, arguments for the “conscription of wealth” simply no longer have the same moral force they once did.

To probe the issue further, consider the following thought experiment. Over the past several years, Representative Charles Rangel, a Korean War veteran, has proposed legislation to reinstitute the draft. The crux of Rangel’s argument is that “military service should be a shared sacrifice” and that we should “not allow some to stay behind while other people’s children do the fighting.”22 The Rangel bill has never passed and,
given strong popular opposition to the draft, it is unlikely to pass anytime soon. But imagine for the moment if the Rangel bill were to pass and Congress began requiring individuals to fight in Iraq against their will. Would Congress in such circumstances enact tax cuts of the EGTRRA/JGTRRA variety? Is it possible to imagine repealing the estate tax or reducing the taxation of capital gains or dividends in an environment where Congress has mandated military service? Perhaps—in politics, one should never say never. However, we submit that debates over how to pay for war are cast in very different terms when soldiers on the frontline include not only those who have volunteered for the assignment but also those who are there under force of law.

Some may regard this as an unfortunate commentary on the politics of war financing in the 21st century. Perhaps arguments for “shared sacrifice” should carry as much political weight when the country’s military efforts are carried out by professional volunteers as when ordinary citizens are drafted into service. Over the past several years, however, there has been little evidence that arguments for shared sacrifice continue to resonate with the American electorate.

* * *

Because the Bush tax cuts represent such a significant departure from the usual wartime practice of raising taxes, commentators have understandably asked whether current policies mark a break from a longstanding patriotic tradition of wartime fiscal sacrifice. Have we entered a new era of fiscal self-indulgence, where even in the face of mounting losses of blood and treasure, American voters demand fewer burdens from their government? As the analysis above suggests, we believe that strands of that mode of thinking about wartime tax policy have surfaced throughout American history.

What is different about the current period is the constellation of circumstances making possible a more extreme manifestation of our nation’s latent instinct to oppose the burdens of taxation. In combination, the three factors described above—historically low inflation rates, a political environment that has marginalized deficit concerns, and the elimination of the draft—have transformed the politics of wartime taxation in the United States. We find it noteworthy that these changes have influenced not only observed policy outcomes (wartime tax cuts rather than wartime tax increases), but also that they have begun to change how we talk about our collective responsibilities during war.

Throughout American history, lawmakers have made the case for higher taxes as an
expression of support for U.S. troops. Indeed, in every conflict we examined, support for higher taxes was viewed as a defining feature of being a “military hawk.” As one GOP senator put it in January 1967, “I just don’t see how we can be hawks on the war and then vote against taxes to pay for it.” There is scant evidence of any remaining life in this point of view. Indeed, at times lawmakers have turned the argument upside down, arguing that only by reducing taxes can we truly support the troops. Speaking in April 2003, for example, Kentucky Senator Jim Bunning made the case for the administration's tax cuts, arguing that “When our troops come home, I hope they have jobs. The Reserves and Guardsmen coming back, their jobs are on the line.” Senator Bunning’s argument stands in stark contrast to the political rhetoric of a half-century earlier, when House Speaker Sam Rayburn admonished his colleagues by noting, “I think the boys in Korea would appreciate it more if we in this country were to pay our own way instead of leaving it for them to pay when they get back.”

It is of course impossible to know how events will unfold over the next several months and years. Most commentators view the elections of November 2006 as a repudiation of the administration’s policies in Iraq. The fact that Democrats now control both chambers of Congress will no doubt affect the future direction of the U.S. military’s role in that country, as will future changes in the White House, especially if a Democrat wins the presidency in November 2008. Even so, it is worth remembering that, with regard to the war financing question, Democrats have so far shown little interest in reversing the administration’s simultaneous pursuit of war and tax cuts. Indeed, if anything, Democrats seem intent on introducing their own brand of tax cuts, even as the war in Iraq continues. If this happens, it might signal that wartime tax cuts, which so many commentators initially decried as a historical anomaly, have found a more secure footing in American politics.
Endnotes

† Excerpted from STEVEN A. BANK, KIRK J. STARK, & JOSEPH J. THORNDIKE, WAR AND TAXES (Urban Institute Press 2008).


4 For an overview and description of these most significant changes, see DAVID L. BRUMBAUGH, CONGRESSIONAL RESEARCH SERVICE, MAJOR TAX ISSUES IN THE 109TH CONGRESS (2005).


11 During World War II, federal spending on national defense averaged 76.4 percent of total outlays, absorbing 37.8 percent of gross domestic product in the final years of the war.


13 KARLYN BOWMAN, AMERICAN ENTERPRISE INSTITUTE, PUBLIC OPINION ON TAXES 14 (2007).


18 Andrew Taylor, *The Ideological Roots of Deficit Reduction Policy*, 19 REV. POL. RES. 11, 26 (2002) (observing that “moderates of both parties are the most vocal advocates of balanced budgets today.”).


20 64 Cong. Rec. 2294 (1917).


