

# **UCLA**

## **Journal of Scholarly Perspectives**

### **Title**

Full Issue

### **Permalink**

<https://escholarship.org/uc/item/7792z0h9>

### **Journal**

Journal of Scholarly Perspectives, 8(01)

### **Author**

UCLA, Law School

### **Publication Date**

2012

### **Copyright Information**

Copyright 2012 by the author(s). All rights reserved unless otherwise indicated. Contact the author(s) for any necessary permissions. Learn more at <https://escholarship.org/terms>

# SCHOLARLY PERSPECTIVES

UCLA | SCHOOL OF LAW

# SCHOLARLY PERSPECTIVES

UCLA | SCHOOL OF LAW

Fall 2012

Volume VIII, No. 1

Rachel F. Moran  
Dean and Michael J. Connell Distinguished Professor of Law

Editors           Lauri L. Gavel  
Executive Director of Communications

Sara Rouche  
Communications Officer

Citation Editor   Cheryl Kelly Fischer  
Reference Librarian

Design           Francisco A. Lopez  
Manager of Publications and Graphic Design

Photography      Todd Cheney  
ASUCLA Photography

## TABLE OF CONTENTS

---

CORPORATE LAWYERS AS GATEKEEPERS 5

Stephen M. Bainbridge

UNDOCUMENTED CRIMINAL PROCEDURE 25

Devon W. Carbado and Cheryl I. Harris

UNDERSTANDING LAW AND RACE AS  
MUTUALLY CONSTITUTIVE 47

Laura E. Gómez

UNDERSTANDING THE RAND COMMITMENT 67

Douglas Lichtman

SUPPORTING WORKERS BY ACCOUNTING  
FOR CARE 89

Noah D. Zatz





## Stephen M. Bainbridge

William D. Warren Distinguished Professor of Law

Stephen Bainbridge currently teaches Business Associations, Advanced Corporation Law and Corporate Governance. In past years, he has taught Corporate Finance, Securities Regulation, Mergers and Acquisitions and Unincorporated Business Associations. In 2008, he received the UCLA Law Rutter Award for Excellence in Teaching. In both 2008 and 2011, he was listed among the 100 most influential people in the field of corporate governance by *Directorship* magazine.

A prolific scholar, his work emphasizes the law and economics of public corporations. He has written more than 75 law review articles for a number of leading law journals. His most recent books are *Corporate Governance after the Financial Crisis* and *Agency, Partnerships and Limited Liability Entities: Cases and Materials on Unincorporated Business Associations, 3<sup>rd</sup> Ed.* (with William Klein and Mark Ramseyer).

Professor Bainbridge graduated from the University of Virginia School of Law, and from 1994 to 1996, served as a Salvatori Fellow with the Heritage Foundation. Before coming to UCLA in 1997, he taught at the University of Illinois Law School; at Harvard Law School as the Joseph Flom Visiting Professor of Law and Business; at La Trobe University in Melbourne, Australia; and at Aoyama Gakuin University in Tokyo. He currently serves on the American Bar Association's Committee on Corporate Laws and on the Editorial Advisory Board of the *Journal of Markets and Morality*. He chairs the Executive Committee of the Federalist Society's Corporations, Securities & Antitrust Practice Group.



# CORPORATE LAWYERS AS GATEKEEPERS\*

Stephen M. Bainbridge

## INTRODUCTION

The capital markets for corporate securities suffer from an inherent information asymmetry. Investors demand credible information both at the time of the original purchase from the issuer and on an on-going basis so as to support a liquid secondary trading market. In theory, corporations will provide such information. In practice, of course, some will choose fraud, and various market failures will cause even honestly run firms to disclose inaccurate or incomplete information. Investors will demand to be compensated for bearing this risk via a higher rate of return.

As a way of both improving the quality of and bonding the credibility of its disclosures, so as to reduce its cost of capital, a company will hire various outsiders—such as an outside auditor and underwriters—to function as reputational intermediaries. Because the gatekeeper’s business depends on its reputation for honesty, probity, and accuracy, it will not ruin that reputation to aid one client to cheat. These outsiders thereby function as gatekeepers, policing access to the capital markets.

Until passage of the Sarbanes-Oxley Act (SOX), lawyers rejected the idea that they were gatekeepers. The corporate bar long insisted that it owed no duties to anyone other than the managers and boards of directors of its clients. The idea that lawyers might have obligations to shareholders, the investing public, or other capital market participants was abhorrent to the bar. Lawyers were advocates, confidants, and advisors, not auditors.<sup>1</sup>

In fact, however, lawyers often play a reputational intermediary role not dissimilar to that of an auditor. A very high profile general counsel or law firm partner, for example, can give a client in trouble the benefit of the lawyer’s reputation for probity and upstanding ethics.

Usually, of course, counsel play a more behind-the-scenes role, but it is still a gate-keeping role. Specifically, transactional counsel and in-house lawyers are well positioned to intervene by blocking the effectiveness of a defective registration statement or prevent the consummation of a transaction, to cite but two examples.<sup>2</sup>



Unfortunately, lawyers have all too often failed to be effective gatekeepers. In litigation arising out of the 1980s savings and loan crisis, for example, Judge Stanley Sporkin famously asked:

Where were these professionals, a number of whom are now asserting their rights under the Fifth Amendment, when these clearly improper transactions were being consummated? Why didn't any of them speak up or disassociate themselves from the transactions? Where also were the outside accountants and attorneys when these transactions were effectuated? What is difficult to understand is that with all the professional talent involved (both accounting and legal), why at least one professional would not have blown the whistle to stop the overreaching that took place in this case.<sup>3</sup>

A decade later, the same questions were asked of lawyers who worked for firms like Enron.

At Enron itself, for example, there was "an absence of forceful and effective oversight [of the company's disclosures] by . . . in-house counsel, and objective and critical professional advice by outside counsel at Vinson & Elkins,"<sup>4</sup> along with senior management and the auditors. The report expressly criticized Vinson & Elkins, which the investigators argued "should have brought a stronger, more objective and more critical voice to the disclosure process."<sup>5</sup>

An internal investigation at WorldCom likewise faulted, among others, the firm's lawyers for allowing a pervasive "breakdown in ... the company's corporate-governance structure."<sup>6</sup> An internal investigation criticized WorldCom's general counsel because his legal department was not properly structured "to maximize its effectiveness as a control structure upon which the Board could depend."<sup>7</sup>

As Senator John Edwards aptly summarized the problem, when companies break the law, "you can be sure that part of the problem is that the lawyers ... are not doing their jobs."<sup>8</sup> Edwards successfully persuaded Congress that "corporate lawyers should not be left to regulate themselves [any] more than accountants should be left to regulate themselves."<sup>9</sup> SOX therefore required the Securities and Exchange Commission (SEC) to adopt new ethics rules bringing the corporate lawyer – client relationship into the federal regulatory sphere.

Despite SOX's many strictures in this and other areas, however, a new and even more devastating financial crisis came in 2008 when the subprime mortgage market's troubles nearly brought the entire banking system to its knees. Once again, questions are being asked about the role that lawyers played in this crisis. A reassessment of SOX's legal ethics rules thus is in order.

The relationship between corporate lawyers and corporate management presents two distinct sets of principal-agent problems. The first arises out of the information asymmetry between full-time managers and independent members of the board of directors who devote but a small portion of their time and effort to the firm. The former inherently have better access to firm information than do the latter. That asymmetry became even more acute, however, with the post-SOX emphasis on having a board whose majority consists of outsiders so insulated from management as to satisfy the demanding definition of independence. The post-reform board therefore must find new ways of unbiased and independent information.

Corporate counsel could be an important source of such information. Because the management-attorney relationship tends to dominate the attorney's relationship with the firm however, lawyers have strong incentives to help management control the flow of information to the board of directors. As Enron Bankruptcy Examiner Neal Batson observed, for example:

One explanation for the attorneys' failure may be that they lost sight of the fact that the corporation was their client. It appears that some of these attorneys considered the officers to be their clients when, in fact, the attorneys owed duties to Enron.<sup>10</sup>

Indeed, as Senator Edwards noted, counsel often develop a de facto loyalty to management that trumps their de jure duties:

We have seen corporate lawyers sometimes forget who their client is. What happens is their day-to-day conduct is with the CEO or the chief financial officer because those are the individuals responsible for hiring them. So as a result, that is with whom they have a relationship. When they go to lunch with their client, the corporation, they are usually going to lunch with the CEO or the chief financial officer. When they get phone calls, they are usually returning calls to the CEO or the chief financial officer.<sup>11</sup>

This problem is especially pronounced for in-house counsel. Technically, of course, they work for the entity, but, in practice, counsel naturally tend to view their management supervisors as their employer.<sup>12</sup>

The second principal-agent problem arises because attorneys may be tempted to turn a blind eye to managerial misconduct or even to facilitate such misconduct. As to in-house general counsel, even if formally appointed by the board of directors, their tenure normally depends on their relationship with the CEO. As for outside

legal counsel, they must please their clients in order to retain their business and to attract the business of future clients. This pressure is especially strong given the large number of capable firms and attorneys available for hire: law firms are something of a fungible good.

In particular, counsel are subject to strong pressure to be a team player. In *The Terrible Truth About Lawyers*, Mark McCormack, founder of the International Management Group, a major sports and entertainment agency, wrote that “it’s the lawyers who: (1) gum up the works; (2) get people mad at each other; (3) make business procedures much more expensive than they need to be; and (4) now and then deep-six what had seemed a perfectly workable arrangement.”<sup>13</sup> McCormack further observed that, “when lawyers try to horn in on the business aspects of a deal, the practical result is usually confusion and wasted time.”<sup>14</sup> He concluded that often “the best way to deal with lawyers is not to deal with them at all.”<sup>15</sup>

Because these attitudes are widely shared in the business community, there is much pressure—especially on in-house counsel—to get out of the way. In turn, lawyers want to be seen as team players. Unfortunately, the incentive to be a team player has led some counsel to bless highly suspect management decisions.

As Enron examiner Batson observed, for example, Enron’s “attorneys saw their role in very narrow terms, as an implementer, not a counselor. That is, rather than conscientiously raising known issues for further analysis by a more senior officer or the Enron Board or refusing to participate in transactions that raised such issues, these lawyers seemed to focus only on how to address a narrow question or simply to implement a decision (or document a transaction).”<sup>16</sup>

To be clear, the point is not that lawyers are pervasively co-opted or immoral. The point is only that lawyers have both economic incentives and cognitive biases that systematically incline them to at least shut their eyes to instances of client misconduct.<sup>17</sup>

## II. SOX § 307

**W**hen the Senate took up SOX, Senator Edwards proposed a floor amendment, subsequently enacted as § 307 of the Act, requiring the SEC to:

[I]ssue rules ... setting forth minimum standards of professional conduct for attorneys appearing and practicing before the Commission in any way in the representation of issuers, including a rule (1) requiring an attorney to report evidence of a material violation of securities law or breach of fiduciary duty or

similar violation by the company or any agent thereof, to the chief legal counsel or the chief executive officer of the company (or the equivalent thereof); and (2) if the counsel or officer does not appropriately respond to the evidence ... requiring the attorney to report the evidence to the audit committee of the board of directors of the issuer or to another committee of the board of directors comprised solely of directors not employed ... by the issuer, or to the board of directors.<sup>18</sup>

It gives lawyers a very “simple” obligation: “You report the violation. If the violation isn’t addressed properly, then you go to the board.”<sup>19</sup>

In compliance with § 307, the SEC, in January 2003, promulgated the “Part 205” attorney conduct regulation.<sup>20</sup> The core of the new rules is a version of the up-the-ladder reporting requirement envisioned by Senator Edwards.<sup>21</sup>

## A. SEC Implementation

The initial jurisdictional question is whether a lawyer is “appearing and practicing before the Commission in the representation of an issuer.”<sup>22</sup> Only lawyers doing so are subject to the SEC’s ethics standards. Unfortunately, the definition of “appearing and practicing” is both sweeping and quite vague:

### 1. Preliminary Matters: Jurisdiction and Preemption

Appearing and practicing before the Commission: (1) Means: ... Providing advice in respect of the United States securities laws or the Commission’s rules or regulations thereunder regarding any document that the attorney has notice will be filed with or submitted to, or incorporated into any document that will be filed with or submitted to, the Commission, including the provision of such advice in the context of preparing, or participating in the preparation of, any such document ....<sup>23</sup>

To be sure, the adopting release states “an attorney’s preparation of a document (such as a contract) which he or she never intended or had notice would be submitted to the Commission, or incorporated into a document submitted to the Commission, but which subsequently is submitted to the Commission as an exhibit to or in connection with a filing, does not constitute ‘appearing and practicing’ before the Commission.”<sup>24</sup> Yet, many non-securities lawyers may know that their documents will be so filed and thus will find themselves “appearing and practicing” before the Commission despite having had no intention of doing so.

The Part 205 regulations recognize that the attorney “represents the issuer as an entity rather than the officers.”<sup>25</sup> As originally proposed, Part 205.3 further provided that an attorney “shall act in the best interest of the issuer and its shareholders.”<sup>26</sup> As finally adopted, however, the relevant rule provides only that “[a]n attorney appearing and practicing before the Commission in the representation of an issuer owes

### 2. The Issuer as Client

his or her professional and ethical duties to the issuer as an organization.”<sup>27</sup> As UCLA law professor Sung Hui Kim notes, the final rules thus “fail to address the situational pressures” that lead counsel to treat the firm’s managers as their real client.<sup>28</sup>

### 3. Up-the-Ladder Reporting

Former ABA Model Rule 1.13 acknowledged the potential need for an attorney to report on suspected wrongdoing within the organization, but it also limited the ability of an attorney to do so effectively. The language of the Rule was discretionary rather than prescriptive, allowing an attorney to use his judgment about whether or not to proceed with reporting evidence of misconduct to the board of directors or even to high-level corporate officers. In contrast, Part 205 uses the prescriptive word “shall” to describe an attorney’s duty. In pertinent part, the rule provides:

If an attorney, appearing and practicing before the Commission in the representation of an issuer, becomes aware of evidence of a material violation by the issuer or by any officer, director, employee, or agent of the issuer, the attorney shall report such evidence to the issuer’s chief legal officer (or the equivalent thereof) or to both the issuer’s chief legal officer and its chief executive officer (or the equivalents thereof) forthwith.<sup>29</sup>

As a result, an attorney will not have the luxury of using his own judgment about whether or not to report wrongdoing once the statutory level of evidence is triggered. As Senator Edwards anticipated, counsel must report up within the chain of command.<sup>30</sup>

The initial obligation of a lawyer who “becomes aware of evidence of a material violation by the issuer or by any officer, director, employee, or agent of the issuer” is to report such evidence to the issuer’s chief legal or executive officer.<sup>31</sup> Subject to a slew of exceptions and alternatives, unless the lawyer “reasonably believes that [that officer] has provided an appropriate response within a reasonable time, the attorney shall report the evidence of a material violation to” the audit committee of the board of directors.<sup>32</sup>

### B. Assessment

The market for legal services gives management a set of carrots by which to align the interests of corporate counsel with their own. SOX did not attempt to change those incentives. Instead, it gave the SEC a set of sticks by which to enlist corporate counsel in preventing fraud and empowering boards of directors. The goal is laudable. But will it work?

#### 1. Relying on Self-Policing Lawyers

In adopting the final rules, the SEC abandoned an initial effort to “exclude the subjective element” from the concept of “reasonable belief.”<sup>33</sup> An attorney who receives what he “reasonably believes is an appropriate and timely response” from management, for example, “need do nothing more.”<sup>34</sup> As a result, the decision to report up the ladder is largely in the hands of the lawyer.

A related problem, having the same effect, is that several key provisions are expressly permissive rather than mandatory. For example, § 205.3(b) characterizes reporting to the board as a “last resort” rather than requiring automatic disclosure of all evidence of wrongdoing to the board. In practice, only a fraction of reports, therefore, will ever make it past the CEO or CLO to the board.<sup>35</sup>

Much the same problem is presented by the purportedly objective standard requiring a lawyer to report “evidence” of misconduct. After an attorney reports evidence of a material violation to management,<sup>36</sup> the manager shall “cause such inquiry into the evidence of a material violation as he or she reasonably believes is appropriate to determine whether the material violation described in the report has occurred ....”<sup>37</sup> The discretionary standard by which lawyers are to determine whether there is evidence a violation has occurred—namely, whether a “prudent attorney” would think it was “reasonably likely” that a material violation had occurred<sup>38</sup>—allows for professional concerns and other conflicts of interest to skew the lawyer’s assessment, minimizing the chances that the potential violation would be reported to management, let alone to the board. Because of the two-tiered reporting system in which disclosure to the board is contemplated only after management inaction, the corporate managers may often look at evidence presented by a concerned attorney, “reasonably” determine that in fact no violation has occurred (notwithstanding ever-present conflicts of interest), and dismiss the whole matter without any knowledge by the board. The attorney would have complied with his statutory obligation to report up the ladder, yet the board’s monitoring function would have been eviscerated.

In addition, research in behavioral economics suggests certain basic cognitive biases that are likely to discourage lawyers from detecting or acting upon management misconduct. Behavioral economists have identified a number of well-documented cognitive errors relevant to the problem in hand. One is the overconfidence bias, which has been defined as “the belief that good things are more likely than average to happen to us and bad things are less likely than average to happen to us.”<sup>39</sup> If a lawyer is subject to this bias, his judgment will be skewed against believing that his clients are bad people committing fraud. A closely related bias is the confirmatory bias, which is defined as the tendency for actors to interpret information in ways that serve their self-interest or preconceived notions. Lawyers who made the decision to associate with a particular firm, therefore, are less likely to recognize management misconduct, because evidence thereof would be inconsistent both with the lawyer’s self-interest in maintaining a relationship with the co-worker and the lawyer’s self-image as someone who identifies and associates with honest people. Taken together, these systematic, decision-making biases generate a type of “cognitive conservatism” that makes a lawyer “likely to dismiss as unimportant or aberrational the first few negative bits of information that she receives regarding the client or situation.”<sup>40</sup>

## 2. The Lawyer Who Cried Wolf

As we saw in the preceding section, the reliance on self-policing may lead to under-reporting. Given the uncertainty about the scope of the rules and the potentially severe sanctions for guessing wrong, however, it seems equally plausible—and equally problematic—that lawyers may err in the other direction. First, invoking the up-the-ladder reporting right to report evidence about possible wrongdoing allows attorneys to cover themselves. When lawyers routinely use reporting up the ladder as a “CYA” mechanism, however, their conduct changes the signaling effect of disclosure. If reporting up the ladder occurs frequently, it will become a routine procedure that does not necessarily indicate real doubt by the attorney about the propriety of the managers’ actions. Indeed, so as to preserve their relationship with management, lawyers may try to take the sting out of disclosing possible misconduct within an organization by de-stigmatizing the practice. Once reporting up the ladder loses its sting, however, the impact of the disclosure is lessened. In such an environment, senior management and the board will not take accusations as seriously as they should.

## 3. Structural Concerns in the Market for Legal Services

The roles played by Vinson & Elkins as Enron’s principal outside counsel can be separated into three distinct categories: first, the aggressive structuring of the controversial special purpose entity transactions used in Enron’s accounting scam; second, drafting Enron’s disclosure documents; and third, conducting an internal investigation of a whistleblower’s allegations.<sup>41</sup> The latter category is a relatively rare undertaking that differs significantly from the far more common transactional work of the first two types. Oddly, however, SOX § 307 and the Part 205 regulations seem better designed to deal with the third context than with either of the first two. It is only in the third context, for example, that lawyers deliberately set out to look for evidence of wrongdoing.

In transactional work, § 307 issues may arise in one of three main ways:

First, counsel may be aware of aggressive or risky conduct by management but is unaware of fraud or other illegality. Vinson & Elkins, for example, most likely “knew of aggressive and risky transactions and reporting decisions [by Enron’s management] but did not have actual knowledge of illegal conduct.”<sup>42</sup> Section 307 does not require counsel to report evidence of uninformed, excessively aggressive, or unethical conduct to the board; however, counsel only must report evidence of fraud and breaches of fiduciary duty. Because much (probably most) of the board’s monitoring function involves preventing (or at least supervising) overly aggressive management, § 307 thus fails to address the basic information asymmetry between management and the board.

Second, counsel may have actively participated in—or at least facilitated—actual fraud. In these cases, the lawyer also is unlikely to report up the ladder, albeit for the different reason that he now has something to hide.

Third, counsel may have grounds for suspicion—but no direct evidence—of fraud or other illegality. In theory, this category presents the best case for an up-the-ladder reporting requirement to successfully aid the board in overseeing management. In practice, however, it is likely to be very rare. Corporate managers are highly unlikely to seek legal assistance with outright fraud, as opposed to conduct that merely pushes the edge of the envelope. In the post-SOX § 307 environment, managers are even more likely to conceal any hint of impropriety from counsel.<sup>43</sup>

As many commentators on the Part 205 regulations complained, the rules may have a chilling effect on attorney–client communication.<sup>44</sup> Even corporate managers not engaged in actual misconduct will not welcome the investigation that an attorney’s reporting up would engender, especially where there is a possibility that counsel will go over their heads. Managers therefore may withhold information from counsel, so as to withhold it from the board, especially when the managers are knowingly pursuing an aggressive course of conduct. Indeed, in many of the recent corporate scandals, the misconduct was committed by a small group of senior managers who took considerable pains to conceal their actions from outside advisors, such as legal counsel.<sup>45</sup> Many commentators complained that § 307 will diminish the quality of the attorney’s representation of the client because counsel will lack unfettered access to information.<sup>46</sup> More pertinent for our purposes, however, the likelihood that an attorney will encounter evidence of misconduct also is reduced.<sup>47</sup>

#### 4. Impact on Lawyer–Client Communication

Counsel, therefore, is most likely to come across evidence of misconduct when conducting an affirmative investigation, such as when performing due diligence in connection with the issuance of securities. Yet, it may be doubted that due diligence often turns up direct evidence of misconduct. In the first place, even a full-fledged accounting audit is not a true forensic audit designed to uncover wrongdoing, but rather only a sampling audit that may entirely miss the problem.<sup>48</sup> In the second place, due diligence is time-consuming. It is therefore expensive. It therefore tends to be done by young associates. As a result, much client misconduct will go undetected by outside counsel because the lawyers with the most direct exposure to the raw data frequently lack experience. In the third place, due diligence is currently limited to issuances of securities. Routine disclosures and other matters constituting “appearing and practicing” before the SEC traditionally have not triggered a due diligence investigation.

Dotcom era frauds typically involved cooking the books so as to raise—or at least support—the firm’s stock price so that the managers could profit from their stock options. The problem is that generally accepted accounting principles (GAAP) provide substantial flexibility, which permits the phenomenon of earnings management by which corporate managers manipulate financial data so that operating results

#### 5. Will Lawyers Know It When They See It?



conform with forecasts. Even trained corporate lawyers often lack the mathematical skills and accounting knowledge to tell the difference between earnings management allowed by GAAP and illegal financial chicanery. In Enron itself, for example, “Enron and its accountants were (in many cases) making exquisitely fine judgment calls.”<sup>49</sup> Few lawyers likely have the expertise necessary to second-guess such judgments. As Professor Lawrence Cunningham observed, “an important lesson from Enron is the danger that prevailing professional cultures create a crack between law and accounting that resolute fraud artists exploit.”<sup>50</sup>

### III. WHAT COULD HAVE BEEN DONE

The SEC was quite reticent in exercising its authority under § 307. Consistent with the spirit of SOX, the SEC might have been even more aggressive in pressing lawyers to communicate with the board of directors. The SEC might have required, for example, that the audit committee and/or the board meet periodically with the general counsel outside the presence of other managers and inside directors.

The SEC might have required the counsel to report possible violations to the board even if the chief legal or executive officer undertook a reasonable response to the violation. After all, it is the board of directors that has a responsibility to assure that “appropriate information and reporting systems are established by management” and that “appropriate information will come to its attention in a timely manner as a matter of ordinary operations.”<sup>51</sup>

A more radical solution would be an enhanced due diligence obligation, which would effectively transform securities lawyers into auditors. A legal audit of the firm in connection with major transactions and/or the preparation of significant disclosure documents would increase the likelihood that counsel would become aware of evidence of client misconduct, which could then be reported up the ladder. Indeed, SOX had already moved in this direction by imposing a new obligation for the chief executive officer and chief and financial officers to certify disclosure documents.<sup>52</sup> Given the amount of client misconduct that went undetected by accounting audits and legal due diligence, however, it may be doubted whether the benefits of such a radical solution would outweigh the costs.

To be sure, these ideas push the edge of the envelope insofar as the SEC’s regulatory authority is concerned. Although § 307 only explicitly mandated an up-the-ladder reporting requirement, the statutory reference to “minimum standards of professional conduct” sweeps far more broadly and could easily encompass additional, more extensive obligations. The SEC thus doubtless has wider authority than it has chosen to exercise to date.<sup>53</sup>

At the end of the day, § 307 – warts and all – was necessary to break the organized bar’s resistance to legal ethics reforms intended to reduce the managerialist bias of the rules of professional conduct. Corporate counsel work for the board, not management. Only by threatening lawyers who fail to report up-the-ladder with discipline could the balance of power be shifted in favor of directors relative to managers.

In practice, firms should still be able to ensure that the client gets the full benefit of transactional lawyering services by developing a best practice approach to dealing with possible material violations. In consultation with the audit committee, the general counsel and principal outside counsel should develop a written policy for identifying and reporting violations. The board members, CEO, and CFO should be briefed on their legal obligations with respect to reports, but also encouraged to view a report as a potential win-win situation rather than a zero-sum or adversarial game. Up-the-ladder reporting can give the firm an opportunity to cut off potential violations before they mature into a legal or public relations nightmare, but only if counsel and managers are willing to trust one another.



- \* This essay is an abridged version of Stephen M. Bainbridge's *CORPORATE GOVERNANCE AFTER THE FINANCIAL CRISIS*, Oxford University Press (2011).
1. See John C. Coffee, Jr., *Understanding Enron: "It's About the Gatekeepers, Stupid,"* 57 *BUS. LAW.* 1403, 1405 (2002) (noting prevailing view that an issuer's lawyer is a "transaction engineer, rather than [a] reputational intermediary").
  2. *SEC v. Nat'l Student Mktg. Corp.*, 457 F. Supp. 682 (D.D.C. 1978), provides a well-known example of how transactional lawyers failed as gatekeepers. National Student Marketing was to be acquired by Interstate National Corporation in a merger. The merger agreement conditioned the parties' obligation to close on an exchange of opinion letters from each corporation's attorneys and "cold comfort letters" from each corporation's accountants. At the closing, National Student Marketing's lawyers became aware of serious problems with their client's financial statements. The problems were sufficiently severe that National Student Marketing's auditors refused to issue a clean cold comfort letter. Despite knowing of both the auditor's concerns and the underlying problems, counsel nevertheless issued the necessary opinion letter and failed otherwise to object to the closing going forward. Whether the lawyers should have blown the whistle on their client remains a matter of debate. At a minimum, however, National Student Marketing's lawyers could and should have closed the gate, preventing the merger from being consummated, by withholding their opinion letters and advising their clients to hold off until the accounting concerns could be addressed.
  3. *Lincoln Sav. & Loan Ass'n v. Wall*, 743 F. Supp. 901, 920 (D.D.C. 1990).
  4. William C. Powers, Jr., et al., Report of Investigation by the Special Investigative Committee of the Board of Directors of Enron Corp. 17 (Feb. 1, 2002), available at <http://news.findlaw.com/hdocs/docs/enron/sicreport/>.
  5. *Id.* at 26.
  6. Rebecca Blumenstein & Susan Pulliam, *WorldCom Fraud was Widespread*, *WALL ST. J.*, June 10, 2003, at A3.
  7. Rebecca Blumenstein & Jesse Drucker, *MCI's Treasurer; Counsel to Resign After Disclosure*, *WALL ST. J.*, June 11, 2003, at A3, A12.
  8. 148 *CONG. REC.* 12519 (2002) (statement of Sen. John Edwards).
  9. *Id.*
  10. See Thomas G. Bost, *Corporate Lawyers After the Big Quake: The Conceptual Fault Line in the Professional Duty of Confidentiality*, 1 *J. BUS. ENTREPRENEURSHIP & L.* 335, 347 (2008) (quoting Batson).
  11. 148 *CONG. REC.* 12519 (2002) (statement of Sen. John Edwards).
  12. Hierarchical pressures are especially pronounced if the managers with whom counsel work with closely have the power to fire them. After all, inside counsel are necessarily economically dependent on a single client. If they get fired, they lose their income, their insurance and other benefits, and their basic livelihood. It is especially difficult for in-house counsel to remain independent of management when a large chunk of their compensation comes in the form of incentive stock

options or restricted stock grants in shares of their employer or when a big chunk of their savings is squirreled away in a 401(k) stuffed with that employer's stock. Indeed, at that point, counsel become subject to the same pressures as any other manager to keep the stock price up, to make sure the company hits its numbers, to puff up good news and downplay bad news. In other words, the in house counsel inherently is subject to a conflict of interest created by dual roles as gatekeeper and as a business person with an interest in the financial success and longevity of the entity.

13. MARK H. McCORMACK, *THE TERRIBLE TRUTH ABOUT LAWYERS* 14–15 (1987).
14. *Id.* at 87.
15. *Id.* at 15.
16. Thomas G. Bost, *The Lawyer as Truth-Teller: Lessons From Enron*, 32 PEPP. L. REV. 505, 514 (2005) (quoting Batson).
17. As my friend and UCLA School of Law colleague Sung Hui Kim noted in an important contribution to the literature, conventional wisdom assumes that lawyers who participate in or otherwise facilitate corporate fraud have “consciously and deliberately chose[n] the path of greed and depravity.” Sung Hui Kim, *The Banality of Fraud: Re-situating the Inside Counsel as Gatekeeper*, 74 FORDHAM L. REV. 983, 991 (2005). As an alternative to this “venality hypothesis,” Kim advances a “banality hypothesis.” *Id.* at 992. According to the latter, inside counsel who behave unethically do so not because they are venal but because their situation causes them to be conformist and obedient. Kim supports her argument by drawing on research from several related disciplines—which might variously be described as behavioral economics, experimental economics, experimental psychology, and sociology—to provide a detailed “ethical ecology” of the environment within which inside counsel operate. *Id.* at 1001. This ethical environment has three critical characteristics: (1) the inside counsel is conceptualized by both himself and others as an employee; (2) employees are professionalized to view themselves as agents whose conduct should be aligned to the interests of the principal; and (3) they are expected to be team players. While arguing that the inside counsel’s situation “makes unethical behavior, at least in the form of acquiescence, likely,” Kim acknowledges that one might expect “smart professional[s]” to nevertheless exercise their moral autonomy and take an “ethical stand.” *Id.* at 1026. In response, Kim deploys theories about systematic biases from cognitive psychology to argue that such professionals are often able to blind themselves to the ethical problem at hand.
18. Sarbanes-Oxley Act of 2002 § 307, 15 U.S.C. § 7245 (2006).
19. 148 CONG. REC. 12519 (2002) (statement of Sen. John Edwards).
20. Implementation of Standards of Professional Conduct for Attorneys, Securities Act Release No. 8185, Exchange Act Release No. 47,276, Investment Company Act Release No. 25,919, 68 Fed. Reg. 6296 (Feb. 6, 2003), available at <http://www.sec.gov/rules/final/33-8185.htm> [hereinafter Securities Act Release No. 8185].
21. 17 C.F.R. § 205.3(d)(2) (2011). In promulgating the Part 205 regulations, the SEC postponed action with regard to mandatory noisy withdrawals. The original proposal obligated an attorney whose internal complaints did not receive an adequate mitigating response by the issuer to resign from the corporation and to file a notification with the

SEC explaining the basis for such resignation. This noisy withdrawal rule met with substantial criticism from the bar. As adopted, Part 205 permits, but does not require, an attorney to disclose confidential client information to the SEC under specified conditions, most notably where necessary to prevent “injury to the financial interest or property of the issuer or investors.” *Id.*

22. 17 C.F.R. § 205.1 (2011).
23. 17 C.F.R. § 205.2(a)(1)(iii) (2011).
24. Securities Act Release No. 8185, *supra* note 20.
25. Press Release No. 2002-158, SEC, SEC Proposes Rules to Implement Sarbanes-Oxley Act Provisions Concerning Standards of Professional Conduct for Attorneys (Nov. 6, 2002), *available at* <http://sec.gov/news/press/2002-158.htm>.
26. Securities Act Release No. 8185, *supra* note 20.
27. 17 C.F.R. § 205.3(a) (2011).
28. Kim, *supra* note 17, at 1034.
29. 17 C.F.R. § 205.3(b)(1) (2011).
30. As discussed below, Part 205 facially preempts inconsistent state rules of legal ethics. *See infra* note 53.
31. § 205.3(b)(1).
32. 17 C.F.R. § 205.3(b)(3) (2011).
33. Securities Act Release No. 8185, *supra* note 20.
34. 17 C.F.R. § 205.3(b)(8) (2011).
35. To be sure, § 205.3(b)(4) allows attorneys to bypass management and go directly to the board to report misconduct if it would be futile to report to the CEO or CLO. The problem with that provision may be best illustrated by an example. Suppose attorney Anne suspects wrongdoing by the CEO. Knowing full-well that reporting the CEO’s wrongdoing to the CEO and/or the CLO would be pointless, but considering that the “futility provision” is only discretionary, Anne may choose to talk to the CEO and CLO rather than go over their heads to the Board for fear of alienating those individuals. If there is any risk that Anne is mistaken about the wrongdoing, going over the heads of the CEO and CLO will permanently poison her relationship with them. If Anne opts to first speak to the CEO or CLO, and they give Anne her statutorily mandated assurance that the problem is being investigated and reasonably resolved, Anne has fulfilled her duties. Yet, the CEO and CLO may be allowed to continue their misconduct.
36. The standard is (confusingly) defined as “credible evidence, based upon which it would be unreasonable, under the circumstances, for a prudent and competent attorney not to conclude that it is reasonably likely that a material violation has occurred, is ongoing, or is about to occur.” 17 C.F.R. § 205.2(c) (2011).
37. 17 C.F.R. § 205.3(b)(2) (2011).
38. § 205.2(c).
39. Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CAL. L. REV. 1051, 1091 (2000).
40. Donald C. Langevoort, *Where Were the Lawyers? A Behavioral Inquiry Into Lawyers’ Responsibility for Clients’ Frauds*, 46 VAND. L. REV. 75, 101 (1993).

41. See generally Jill E. Fisch & Kenneth M. Rosen, *Is There a Role for Lawyers in Preventing Future Enrons?*, 48 VILL. L. REV. 1097 (2003).
42. *Id.* at 1115.
43. In addition, lawyers will very rarely perceive their own situation to fall within the third scenario. As former Delaware Chancellor Allen aptly noted, albeit in a rather different context, “human nature may incline even one acting in subjective good faith to rationalize as right that which is merely personally beneficial.” *City Capital Assocs. v. Interco Inc.*, 551 A.2d 787, 796 (Del. Ch. 1988). It typically is personally beneficial for lawyers to refrain from antagonizing the corporate managers who hire and fire them. Hence, absent the proverbial smoking gun, we can expect lawyers to turn a blind eye to indicia of misconduct by those managers. Section 307 does too little to change those incentives.
44. See Securities Act Release No. 8185, *supra* note 20 (summarizing commentators’ views).
45. See, e.g., Mark Maremont, *Rite Aid Case Gives First View of Wave of Fraud on Trial*, WALL ST. J., June 10, 2003, at A1 (describing the “great lengths” to which defendants went in order to “hide their tracks”).
46. See Securities Act Release No. 8185, *supra* note 20 (summarizing commentators’ views).
47. If counsel cannot be bypassed, management may seek to intimidate them. The Enron abuses long went undetected, in part, because Enron management “either overruled or intimidated” subordinates. Steven L. Schwarcz, *Enron and the Use and Abuse of Special Purpose Entities in Corporate Structures*, 70 U. CIN. L. REV. 1309, 1317 (2002). In-house counsel may be particularly vulnerable to such pressures.
48. Larry E. Ribstein, *Market vs. Regulatory Responses to Corporate Fraud: A Critique of the Sarbanes-Oxley Act of 2002*, 28 J. CORP. L. 1, 31 (2002).
49. Schwarcz, *supra* note 47, at 1313.
50. Lawrence A. Cunningham, *Sharing Accounting’s Burden: Business Lawyers in Enron’s Dark Shadows*, 57 BUS. LAW. 1421, 1422 (2002). In the selective disclosure context, the SEC acknowledges that: “In many cases, an issuer’s chief financial officer or investor relations officer may have a keener awareness than company counsel of the significance of information to investors.” Report of Investigation Pursuant to Section 21(a) of the Securities Exchange Act of 1934: Motorola, Inc., Exchange Act Release No. 46,898 (Nov. 25, 2002), available at <http://www.sec.gov/litigation/investreport/34-46898.htm>. In my view, the qualifiers “many” and “may” were unnecessary.
51. *In re Caremark Int’l Inc. Derivative Litig.*, 698 A.2d 959, 969–70 (Del. Ch. 1996).
52. Sarbanes-Oxley Act of 2002 § 302, 15 U.S.C. § 7241 (2006).
53. Such radical reforms also would have conflicted with legal ethics rules of many states. The Part 205 regulations facially preempt state rules of professional conduct, however. 17 C.F.R. §§ 205.6(b)-(c) (2011). Accordingly, where there is conflict between a state’s rules and Part 205, the latter prevails, unless the state imposes a more stringent obligation upon its attorneys that is consistent with Part 205. Attorneys who comply with the Regulation’s procedures in good faith will be immune from liability for violating state ethics rules that conflict. As a result, the organized bar likely would be pressured to square its rules with those promulgated by the SEC.







## Devon W. Carbado

Professor of Law

Devon Carbado teaches Constitutional Criminal Procedure, Constitutional Law, Critical Race Theory and Criminal Adjudication and recently served as Vice Dean of the Faculty. He was elected Professor of the Year by the UCLA School of Law Classes 2000 and 2006 and received the Rutter Award for Excellence in Teaching in 2003 and the Eby Award for the Art of Teaching in 2007. He is a recipient of the Fletcher Foundation Fellowship, which is awarded to scholars whose work furthers the goals of *Brown v. Board of Education*.

He writes in the areas of critical race theory, employment discrimination, criminal procedure, constitutional law and identity. He is editor of *Race Law Stories* (with Rachel Moran) and is working on a book on employment discrimination (with Mitu Gulati). He is a former director of the Critical Race Studies Program at UCLA Law, a faculty associate of the Ralph J. Bunche Center for African American Studies, a board member of the African American Policy Forum and a James Town Fellow.

After graduating from Harvard Law School in 1994, Professor Carbado joined Latham & Watkins in Los Angeles and was subsequently appointed a Faculty Fellow and Visiting Associate Professor of Law at the University of Iowa College of Law.



## Cheryl I. Harris

Rosalinde and Arthur Gilbert Professor of Civil Liberties and Civil Rights

Cheryl I. Harris teaches Constitutional Law, Civil Rights, Employment Discrimination and Critical Race Theory. In 2005 she received the ACLU Foundation of Southern California Distinguished Professor Award for Civil Rights Education.

As the National Co-Chair for the National Conference of Black Lawyers for several years, she developed expertise in international human rights, particularly concerning South Africa. She was a key organizer of major conferences during the development of South Africa's first democratic constitution in 1994. Her publications include the highly influential "Whiteness as Property" (*Harv. L. Rev.*). Her scholarship has also focused on how racial frames shape our understanding and interpretation of events like Hurricane Katrina and admissions policies as well as how these frameworks influence anti-discrimination law itself.

Professor Harris began her teaching career at Chicago-Kent College of Law in 1990 after more than a decade in practice that included criminal appellate and trial work and municipal government representation for the city of Chicago. In 2002, she received a fellowship from the Mellon Foundation to co-host a semester-long interdisciplinary working group and conference series, "Redress in Social Thought, Law and Literature," at the University of California Humanities Research Institute. She is a member of the Advisory Board of the Ralph J. Bunche Center for African American Studies and has served as a consultant to the MacArthur Foundation.



# UNDOCUMENTED CRIMINAL PROCEDURE\*

Devon W. Carbado\*\*

Cheryl I. Harris

For more than two decades, criminal procedure scholars have debated what role, if any, race should play in the context of policing. Although a significant part of this debate has focused on racial profiling, or the practice of employing race as basis for suspicion, criminal procedure scholars have paid little attention to the fact that the U.S. Supreme Court has sanctioned this practice in a number of cases at the intersection of immigration law and criminal procedure. Notwithstanding that these cases raise similar questions to those at the heart of legal and policy debates about racial profiling, they are largely overlooked in the criminal procedure scholarship on race and policing. We refer to these cases as the undocumented cases. While there are a number of doctrinal and conceptual reasons that explain their marginalization, none of these reasons are satisfying given the importance of the undocumented cases to debates about race, racial profiling, and the Fourth Amendment. The undocumented cases import a pernicious aspect of immigration exceptionalism into Fourth Amendment doctrine—namely, that the government can legitimately employ race when it is enforcing immigration laws. In so doing, the cases constitutionalize racial profiling against Latinos and unduly expand governmental power and discretion beyond the borders of immigration enforcement. This weakens the Fourth Amendment and enables racial profiling in the context of ordinary police investigations.

**T**he criminalization of immigration violations and the imposition of immigration sanctions for criminal violations have produced a vexed set of procedural, constitutional, and policy issues.<sup>1</sup> Underlying much of the debate is the generalization (a reasonable one with which we agree) that citizenship affords a set of procedural and constitutional protections in immigration and criminal law enforcement contexts that are unavailable to noncitizens.<sup>2</sup> For example, subsequent to arrest, citizens but not noncitizens are typically eligible for release on bail, while noncitizens but not citizens may be vulnerable to deportation.<sup>3</sup> These and other post-arrest differences highlight the fiction of what Ingrid Eagly calls “doctrinal equality” or the erroneous notion that “noncitizen defendants occupy the same playing field as other defendants in the federal criminal system.”<sup>4</sup>

INTRODUCTION

However, before the arrest, during the investigatory stages of the criminal or immigration process in which law enforcement officials detain, question, and search suspects, the dichotomy between the citizen and noncitizen is, at least in some contexts, decidedly less clear. In particular, the line between citizen and noncitizen is mediated by and bears the racial imprint of a particular historical feature of U.S. immigration law—the government’s explicit employment of race as a proxy for citizenship.<sup>5</sup> In the context of contemporary immigration enforcement, and with respect to Latinos, this proxy function of race blurs the boundary between citizen and noncitizen and further conflates non-citizenship and undocumented status. To make the point concrete, the simple “fact” of apparent Latino ancestry renders a person presumptively an undocumented noncitizen—or, to invoke the unfortunate quasi-term of art, an “illegal alien.”<sup>6</sup> This does not mean that immigration officials and law enforcement personnel actually believe that most or all Latinos are undocumented. The point is that because Latino identity is deemed relevant to the question of whether a person is undocumented, all Latinos live under a condition of presumed illegality.<sup>7</sup>

Fourth Amendment jurisprudence plays a critical role in constructing this problem. The doctrine facilitates both the idea that Latinos are presumptively undocumented (the racial profile)<sup>8</sup> and the practice of detaining Latinos because of that presumption (racial profiling).<sup>9</sup> Yet, for the most part, criminal procedure scholars have not engaged this racial dynamic.<sup>10</sup> They have raised concerns about whether racial profiling creates the crime of “Driving While Black,” about the legitimacy and efficacy of the practice, about whether it imposes a “racial tax”<sup>11</sup> and about whether racial profiling is systemic or derives instead from the actions of a few wayward police officers—the proverbial “bad apples.”<sup>12</sup> But, largely absent from these debates is the fact that law enforcement personnel routinely employ Latino racial identity as a basis for determining whether a person is undocumented or “illegal.”

In a series of cases—*United States v. Brignoni-Ponce*,<sup>13</sup> *INS v. Delgado*,<sup>14</sup> and *United States v. Martinez-Fuerte*<sup>15</sup>—the Supreme Court has sanctioned this practice. While none of these cases explicitly endorse racial profiling per se, each facilitates the practice of utilizing the “appearance of Mexican ancestry” as an investigatory tool. We refer to the cases as the “undocumented cases” because they are marginalized in and mostly omitted from criminal procedure scholarship and casebook discussions about race and the Fourth Amendment.<sup>16</sup> Instead, criminal procedure scholars more frequently engage three other cases to discuss racial profiling: *Terry v. Ohio*,<sup>17</sup> *Whren v. United States*,<sup>18</sup> and *Florida v. Bostick*.<sup>19</sup> We refer to these cases, cumulatively, as the documented cases because they reside within the interior of the criminal procedure literature.

In advancing this argument, we do not mean to suggest that no criminal procedure scholar has engaged *Brignoni-Ponce*, *INS v. Delgado*, or *Martinez-Fuerte*. Our claim is that few criminal procedure scholars invoke these cases as a basis for their racial critiques of Fourth Amendment doctrine.

In the full article upon which this essay is based, we offer a number of reasons why criminal procedure scholars should pay close attention to the undocumented cases. For one thing, these cases broaden our understanding of pretextual seizures by, for example, illustrating the extent to which the government can employ traffic stops as a pretext for enforcing immigration laws. For another, they highlight the importation of a pernicious aspect of immigration exceptionalism into criminal procedure: namely, that the government can legitimately target on the basis of race when it is enforcing immigration laws. Still a third reason scholars should engage the *Delgado*, *Brignoni-Ponce*, and *Martinez-Fuerte* trilogy is that the impact of these cases on Fourth Amendment doctrine transcends the borders of immigration enforcement and shapes the development of Fourth Amendment jurisprudence more generally by both unduly expanding governmental power and facilitating racial profiling.

We will not, in this redacted version of the article, elaborate upon these arguments. Our aim is to focus on why the undocumented cases are marginalized in the criminal procedure canon. The reasons range from how scholars and policymakers frame racial profiling, to how law and social practices racialize Latinos, to how immigration law has doctrinally developed as exceptional, to how Fourth Amendment taxonomies cabin certain searches and seizures as regulatory and administrative and thus not subject to ordinary Fourth Amendment requirements. We contend that none of the explanations either alone or collectively justify the under-examination of the undocumented cases in the criminal procedure literature.

One explanation for the marginalization of the undocumented cases in the criminal procedure scholarship relates to the dominant way in which scholars and policy makers have framed racial profiling. According to some scholars, profiling has its origins in the development of criminal profiles developed in the context of airline hijackings.<sup>20</sup> In the 1970s, this practice subsequently evolved into the creation and use of drug-courier profiles, particularly at airports.<sup>21</sup> At the border, the reliance on race, and at times race alone, as a basis for stopping and investigating travelers was a common practice, and indeed law enforcement operated with relatively few constraints as searches were permitted without warrants or probable cause.<sup>22</sup> However, the extension of investigatory authority to areas outside the border provoked a set of legal challenges to the authority of the U.S. Border Patrol and its practices with regard to immigration enforcement.<sup>23</sup>

**I. EXPLAINING UNDOCUMENTATION: CONCEPTUAL BARRIERS THAT MAKE RACIAL PROFILING IN IMMIGRATION ENFORCEMENT CASES INVISIBLE**

**a. Framing Racial Profiling: Latinos on the Borders**

Policymakers, activists, academics, and lawyers, among others, debated whether immigration officials could employ race (and the focus was on Mexican ancestry) as an immigration enforcement mechanism. At the same time, they debated what evidentiary standard—probable cause, reasonable suspicion (something else?)—would govern immigration enforcement activities in the interior.<sup>24</sup>

This was the context out of which *Brignoni-Ponce* arose in 1975.<sup>25</sup> Yet, ironically, racial profiling was initially framed in criminal procedure scholarship as a problem affecting Blacks, not Latinos. Indeed, the very term racial profiling was initially employed to describe the reliance on racial stereotypes as indicia of criminality primarily vis-à-vis African Americans. The first time “racial profiling” appeared in a law review or law journal was in an article written by Greg Williams, titled “Selective Targeting in Law Enforcement,”<sup>26</sup> that principally focused on Black drivers, although Latino drivers were briefly mentioned.

We tracked the trajectory of literature on racial profiling from 1996, when the term “racial profiling” first appeared, to 2010 by utilizing Westlaw databases. In five-year intervals we compared how the term “racial profiling” appeared in conjunction with “Black/African American,” with “Latino/Hispanic/Mexican,” and with “Arab/Muslim/Middle Eastern/South Asian.” The following table summarizes the results.<sup>27</sup>

Table 1. Frequency of Term “Racial Profiling” by Racial Category

|                                  |           | <i>Racial Profiling</i> | <i>Racial Profiling + Black/African American</i> | <i>Racial Profiling + Latino/Hispanic/Mexican</i> | <i>Racial Profiling + Arab/Muslim/Middle Eastern/South Asian</i> |
|----------------------------------|-----------|-------------------------|--|---|--|
| Date of First Appearance         |           | 1996                    | 1996   | 1996  | 1999   |
| Number of Articles With Terms in | 1995–1999 | 65                      | 40   | 15  | 1  |
|                                  | 2000–2004 | 1298                    | 600  | 293   | 260  |
|                                  | 2005–2009 | 1301                    | 466  | 253   | 245  |
|                                  | 2010      | 206                     | 69   | 42  | 27   |
| Total Frequency*                 |           | 2870                    | 1175   | 603   | 533  |

\* This is the total number of references to race and racial profiling. Included here are references to these terms in conjunction with the specific racial groups mentioned as well as general references to race and racial profiling in which no specific racial group was discussed.

As the table reflects, after the initial period ending in 1999, there was a sharp increase in the number of references to racial profiling from 2000 to 2004—from 65 to almost 1,300. In this interval, Blacks were mentioned with the greatest frequency—600 times, while Latino/Hispanic/Mexican and Arab/Muslim/Middle Eastern/South Asian were mentioned 293 and 260 times respectively. It is also worth noting that the frequency for Latino/Hispanic/Mexican and Arab/Muslim/Middle Eastern/South Asian in this period is roughly the same, though the latter group did not appear until 1999.<sup>28</sup> The subsequent period—from 2005 to 2009—and 2010 show the same pattern: Blacks appear more often in conjunction with the term “racial profiling” than Latinos who appear about as often as Arab/Muslim/Middle Eastern/South Asian. Undoubtedly the events of September 11, 2001, and the intensified focus on Muslims and people of Arab, Middle Eastern, and/or South Asian descent based on the presumption that they pose potential threats to national security help explain the prevalence of these references after 2001. Even as criminal procedure scholarship and analysis of racial profiling clearly increased significantly and incorporated references to other groups, Blacks still appeared to be the major focus whether measured by overall reference to all racial groups or to total times the term appeared.

This is not to say that the focus on how racial profiling impacted Black people and communities was incorrect or inappropriate. To the contrary, Blacks in general and Black motorists in particular were clearly subjected to racially targeted policing practices, some of which were undoubtedly produced by reliance on race-based suspicion.<sup>29</sup> Rather, our point is simply that Latinos were also routinely subjected to racial profiling as part of the effort to combat what was cast as an immigration crisis produced by Mexicans flooding over the border.<sup>30</sup> Simultaneously, Arab/Muslim and Middle Eastern communities were also subjected to racialized suspicion and racial profiling on grounds of national security.<sup>31</sup> Yet legal scholarship on racial profiling focused more on the issue as a Black concern. Apart from several notable exceptions,<sup>32</sup> the criminal procedure literature did not convey the message that racial profiling operated in powerful and pernicious ways across communities and was equally salient for Latinos.

From the outset the paradigmatic case was framed as a traffic stop involving a Black motorist. “Driving While Black” was invoked to articulate a complaint about the use of race as a basis for determining whether a particular driver should be stopped.<sup>33</sup> While Latinos and other racial groups (particularly Muslims, Arabs, Middle Easterners, and South Asians after September 11, 2001) appeared in significant respects over time, the literature tended to reflect and reinforce the notion that racial profiling was a Black experience.



In one respect the problem of the undocumented cases reflects a particular dimension of how Latinos are racially constructed. The erasure or omission of Latinos derives from how they are racialized: Latinos are formally and legally white but socially regarded as nonwhite and inferior. This in-between racial status—"off-white" as Laura Gómez has called it—complicates and often obscures Latino experiences as distinctly racial. This ambivalent racial position of Latinos—sometimes an ethnic group, at other times distinctly a nonwhite race—has historical origins and prior juridical manifestations.<sup>34</sup> Indeed, the "off-white" framing has surfaced in 20th century equality struggles waged by Latinos and persists in contemporary understandings of Latino identity. Even now, the liminal racial position Latinos occupy can sometimes erase their racial experiences or make it difficult to understand those experiences in racial terms. Put another way, Latino racial identity is itself furtive or undocumented.

Against the notion of Latinos as an ethnic group, the historical record reveals a complex and conflicted pattern of racialization. As Gómez has explained, the origins of Latino racialization lie not in immigration and border policing, as these were largely not major concerns until the 1930s and 1940s.<sup>35</sup> Rather the construction of Latino identity is tied specifically to a history of imperialism and conquest in which the United States sought to expand its territory and influence. To begin, the presence of Mexicans in the United States was largely a consequence of the movement of a border rather than the movement of people. The delineation of the border was the result of the U.S. – Mexican war. Launched in the 1840s, this war was justified by assertions of Manifest Destiny and racialized conceptions of national hierarchy that presumed both Anglo-Saxon superiority and nonwhite inferiority.<sup>36</sup> Even the debate over the wisdom of the war and the expansion of the United States into Mexico was deeply influenced by race. The acquisition of the massive Mexican Cession posed both an unprecedented opportunity and a dilemma: What would be the status of the Mexicans living on the acquired land? At a time when citizenship was legally restricted to whites,<sup>37</sup> the incorporation of a group largely perceived as racially inferior posed a vexing issue.<sup>38</sup> Ultimately, under the Treaty of Guadalupe Hidalgo, ending the war in 1848, over 100,000 Mexicans were incorporated into the polity as territorial citizens and thus were de facto "white."<sup>39</sup>

Yet, while Mexican Americans were legally classified as white, they were socially perceived as nonwhite and in some regions were subjected to severe social segregation.<sup>40</sup> Mexican identity was thus distinctly racial, albeit ambiguously positioned below whites and above other nonwhites. In contrast to rules of hypodescent under which any drop of Black blood rendered one Black, a reverse one-drop rule co-evolved in which any drop of Spanish blood made one white, at least in a formalistic sense.<sup>41</sup> Both rules of racial assignment served as technologies of racial

subordination, particularly conferring on early Mexican Americans an incentive to assert their whiteness in order to differentiate and racially distance themselves from Indians and Blacks.

To illustrate the ambivalent, off-white status of Mexicans, consider the naturalization petition of *In re Rodriguez*,<sup>42</sup> decided by a federal judge in Texas in 1897. While the petitioner Ricardo Rodriguez was apparently phenotypically Mexican, was not literate in English (though he testified in English), and averred that he was “pure-blooded Mexican,”<sup>43</sup> he claimed that he was white and thus eligible for U.S. citizenship.<sup>44</sup> The court conceded that Rodriguez was not white in appearance nor would an anthropologist classify him as white.<sup>45</sup> Yet the judge ruled that Rodriguez was “white enough” because of laws that had effectively treated Mexicans as white.<sup>46</sup> These included, in particular, the 1848 Treaty of Guadalupe Hidalgo under which Mexicans living in the Cession were collectively naturalized as citizens of the newly acquired territories. The conflict among whites over whether Mexicans were “white enough” was certainly not definitively resolved by this case. Rather, the dispute reflected the contingent and contested nature of Mexican claims to whiteness.

The legacy of this early history is that the racial dimensions of Latino identity are largely obscured. This difficulty has manifested itself even in the context of efforts to dismantle Jim Crow practices against Mexican Americans. Reflecting the orientation of the first Mexican American civil rights organization, the League of United Latin American Citizens (LULAC), early civil rights advocacy on behalf of Mexican Americans was often framed around the assertion that Mexican Americans were white.<sup>47</sup> The landmark case of *Westminster School District v. Mendez*<sup>48</sup> decided in 1947 is illustrative. The United States Ninth Circuit Court of Appeals ruled that the segregation of Mexican descendant school children in the absence of a state law authorizing the segregation of Mexicans amounted to a denial of equal protection. Notably, while California law did authorize the segregation of children “belonging to one or another of the great races of mankind”—read by the court as Caucasoid, Mongoloid, and Negro—and allowed for segregation of Indians, “Asiatics,” and Blacks, state law did not authorize segregation “within one of the great races.”<sup>49</sup> To the extent then that Mexican Americans were racially identified as whites, they could not be segregated from other whites.

The off-white status of Mexican-Americans figured significantly in *Hernandez v. Texas*,<sup>50</sup> the landmark civil rights case that was decided the same year as *Brown v. Board of Education*.<sup>51</sup> The U.S. Supreme Court held that the systematic exclusion of Mexican Americans from jury duty in a Texas county with significant Mexican American population violated the Fourteenth Amendment. The record established that no Mexican American had served on a jury in 25 years and that even in the

courthouse where the case was tried, men's bathrooms were segregated and marked, "Colored Men and 'Hombres Aqui.'"<sup>52</sup>

Yet neither the parties nor the Court contended that the discrimination involved was race-based. Indeed, the government early on attempted to defend against the charge of discrimination on the grounds that the Latino challengers were white and thus could not establish that they were the victims of racial discrimination at the hands of other whites.<sup>53</sup> The Supreme Court rejected this argument, finding that there was sufficient evidence "to prove that persons of Mexican descent constitute a separate class ..., distinct from 'whites.'"<sup>54</sup> Nevertheless, the court did not hold that this "separate class" constituted a distinct racial group, though their subordinate status was key to the Court's ruling. The ambivalence surrounding Latino racial identity permeated both the way the case was argued and ultimately decided.

Similarly *Perez v. Sharp*,<sup>55</sup> a state constitutional challenge to California's antimiscegenation statute, rested upon the off-white status of Mexicans. The litigation was grounded in the fact that Andrea Perez, a Mexican American woman, was legally classified as white and thus was prohibited from marrying her African American partner. Ruling in favor of Perez, the California Supreme Court rejected the ban on interracial marriage because the state's objective—to preserve racial purity—was inherently impermissible.<sup>56</sup> *Perez* was a major civil rights victory and thus formed an important template for *Loving v. Virginia*,<sup>57</sup> where 20 years later, the Supreme Court struck down similar antimiscegenation laws as unconstitutional under the Fourteenth Amendment. Yet, in the context of the litigation, Latinos were represented in this civil rights dispute as formally white.

All of this is to suggest that the erasure of Latinos as a race has a long juridical history facilitated by their formal classification as white. Conceiving of Latinos as whites—as an ethnic rather than a racial group—has made their experience with racial profiling less legible as a racial practice. Thus, it is perhaps unsurprising that this elision impacts how Latinos are (not) seen in the context of racial profiling debates and the criminal procedure issues racial profiling implicates.

### c. Documenting the Logic of Racial Profiling

One reason why the undocumented cases are marginalized in the criminal procedure debates about racial profiling is that one can view the cases as not in fact authorizing racial profiling, but rather legitimizing the reliance on race when relevant and efficient in enforcing immigration law. Recall that *Brignoni-Ponce* specifically condemns the reliance on racial identity as the sole basis for establishing reasonable suspicion of immigration violations, but permits consideration of race as one of several factors in making that assessment.<sup>58</sup> This maps onto a longstanding debate about what racial profiling actually means.<sup>59</sup> One view holds that racial profiling

refers to the use of race as the *sole* factor in making a decision to investigate, while the other view holds that it includes *any* reliance on race—even reliance on race as one of several factors.<sup>60</sup> On the former view (profiling means relying on race alone) arguably the undocumented cases do not authorize the practice of racial profiling and therefore ought not to be critiqued on that basis. Racial profiling is conceptually defined to exclude what the undocumented cases actually authorize.

Although we have a view as to how scholars and policymakers should conceptualize racial profiling,<sup>61</sup> our intervention transcends this dispute. Whether immigration officials employ race as the sole factor or one of several in establishing reasonable suspicion, they are racially constructing all Latinos as presumptively “illegal”—which is to say, these race-conscious practices participate in constructing the racial category itself.<sup>62</sup> There are several ways to understand this. First, the association between Mexican or Latino racial identity and “illegality” is in part based on conflating the fact that most undocumented people are from Mexico with the erroneous assumption that most Latinos are undocumented.<sup>63</sup> This conflation effectively constructs all Latinos as presumptively undocumented. Second, so long as Latino identity is deemed a relevant factor in determining whether a person is undocumented, all Latinos live under a cloud of this suspicion. The racial category per se is suspicious. Third, the very reason many people are comfortable drawing a racial association between perceived Latino ancestry and “illegal” status is that the racial construction of Latino identity arises from and is bound up with particular narratives about illegality and immigration that are rooted in the past and are now deeply embedded in public discourse and in law. A brief history of immigration enforcement at the border helps to make this plain.

According to historian Kelly Lytle Hernández, Mexican racial identity is specifically rooted in the history of the U.S. Border Patrol, the principal agency charged with enforcing immigration law.<sup>64</sup> The Border Patrol, created in the wake of the passage of the National Origins Act of 1924, was a new federal agency with a broad but ill-defined mandate that left much of its agenda and priorities to be defined through debates between competing social and political forces.<sup>65</sup>

The Border Patrol’s eventual focus on the southern border<sup>66</sup> and the targeting of Mexican immigrant workers beyond the border proper was produced by contestations among nativists, Southwestern agribusinessmen seeking a reliable labor force, the Mexican government’s pursuit of emigration control, and, importantly, the interests of the Border Patrol agents themselves. The latter, as landless members of the working class, were able to create and fulfill a critical role in the borderlands by policing the area’s principal workforce.<sup>67</sup> However, the men of the Border Patrol were not simply engaged in serving the interests of business elites. As members of

the white working class who had been most vehemently opposed to Mexican immigration, they sought to assert authority, achieve power, and in effect buttress their claims to whiteness through their work.<sup>68</sup>

The enforcement of immigration law in the borderlands called upon law enforcement to delineate between the “legal” and “illegal,” and in the context of the highly racialized dynamics of the Southwest, that line was specifically “mexicanized.”<sup>69</sup> Indeed, the primary target of immigration enforcement could be described by Border Patrol agents with some specificity as “Mexican male, about 5’5” to 5’8”; dark brown hair; brown eyes; dark complexion; wearing huaraches . . . and so on.”<sup>70</sup> The institutionalization of the Border Patrol as a federal agency and the process by which it defined its priorities ensured that illegal immigration had a specific racial face. It was a crucible through which “persons guilty of the act of illegal immigration [were transformed] into persons living with the condition of being illegal.”<sup>71</sup> This racial construction, along with the others we have described, ultimately played a critical role in shaping the law of racial profiling.<sup>72</sup>

In this respect, the problem with *Brignoni-Ponce* and the other undocumented cases is not simply that they reflect a faulty empiricism; whether most Latinos are in fact undocumented (they are not) or whether most undocumented people are Latino (they are) is really not the point. The point is that Latino racial identity has been and is presently defined through the trope of illegality such that the racial profile of “illegal alien” is central to their cognizability as a racial group. The undocumented cases reflect and further reinforce this racial recognition, encoding Latino racial identity writ large as a mark of “illegality.”<sup>73</sup>

## II. DOCTRINAL BARRIERS TO UNDOCUMENTATION: THE PLENARY POWER DOCTRINE

In addition to the conceptual barriers outlined above, there are three doctrinal obstacles to the full consideration of the implications of the undocumented cases for core debates about race, policing, and criminal procedure: (1) the plenary power doctrine, (2) the perceived prudential value of *Brignoni-Ponce*, and (3) the regulatory or administrative context of the undocumented cases. These obstacles are discussed in the full version of this article. Below we discuss only the first doctrinal barrier to undocumented cases, the plenary power doctrine.

Although the government has employed immigration law to racially exclude and subordinate nonwhite racial groups, courts have, for the most part, legitimated these practices because of the federal government’s plenary power over immigration—a power that has been said to be absolute.<sup>74</sup> This doctrinal exemption has buttressed a form of immigration exceptionalism that justifies taking even explicitly racial practices, like the use of racial profiles, outside the reach of judicial intervention.

Historically, immigration law has been a site where policymakers, along with the courts, have expressly employed race as a mechanism in determining admission, exclusion, and belonging, and relied on a well-documented history of racist logics to do so. The genesis of this tradition lies in racist opposition to Chinese immigration that was initially welcomed but later vehemently resisted.<sup>75</sup> The Page Act of 1875 marked the beginning of federal immigration regulation and specifically excluded racially marked categories of persons—Chinese prostitutes and “coolie” labor.<sup>76</sup> This was followed by the 1882 Chinese Exclusion Act, which explicitly excluded a racially defined class.<sup>77</sup> Racial hierarchy and the quest for racial homogeneity subsequently shaped the national origins quota system which permitted more immigrants from Western Europe than from Southern or Eastern Europe, Africa, Asia, or Central America.<sup>78</sup>

These racialized practices were naturalized in law and largely treated as justifiable assertions of governmental authority under the doctrine of plenary power. In *Chae Chan Ping v. United States*,<sup>79</sup> the Court ruled that Congress had the power to exclude immigrants of a particular race—here the Chinese—and that this was not subject to judicial review even though the statute contravened lawfully executed treaties between China and the United States and even though Chae held a valid certificate of residency. According to the Court, the enactment of the Chinese Exclusion Act that took effect a week before Chae’s attempt to reenter was within the government’s power to exclude aliens as a necessary aspect of national sovereignty.

Race was central to the Court’s reasoning as the plenary power to protect sovereignty was deemed crucial to protect against a foreign race—a group that “remained strangers in the land,” who could not assimilate and thus were perceived to pose “a great danger that . . . our country would be overrun by them.”<sup>81</sup> Shortly thereafter, this power to exclude was deemed to extend to authorize the deportation of aliens already residing in the United States. In *Fong Yue Ting v. United States*,<sup>82</sup> the Court upheld the Chinese Exclusion Act of 1892, ruling explicitly that the power to regulate immigration and to expel could be exercised on racial grounds.

The plenary power doctrine, then, emerged in a racial context and developed as an explicitly racialized body of law. At the same time, as scholars like Gabriel Chin and Hiroshi Motomura have argued, the doctrine has rendered racialized immigration enforcement remarkably resistant to civil rights intervention.<sup>83</sup> Despite many critiques<sup>84</sup> and even some erosion of plenary power,<sup>85</sup> constitutional principles of equal protection and due process are erratically applied to noncitizens even in the contemporary sphere because of the continued vitality of the plenary power doctrine.<sup>86</sup> While the explicit racial rhetoric of the Chinese Exclusion cases is no longer invoked, what persists is the underlying logic of the cases that linked the

notion of unfettered power to define the terms of admission and exclusion to issues of national sovereignty. To the extent that immigration is deemed to be an arena in which the plenary power doctrine effectively forecloses claims of discriminatory enforcement, the terrain in which Latino racial identity is often defined—immigration enforcement—stands outside of the traditional constraints on and critiques of the use of race in domestic law enforcement. This is the sense in which immigration exceptionalism obscures the significance of the undocumented cases and the issue of racial profiling in immigration enforcement.

## CONCLUSION

Our point of departure for this article was the observation that notwithstanding a rich literature on race, racial profiling, and the Fourth Amendment, criminal procedure scholars have largely ignored the undocumented cases. The marginalization of the undocumented cases in the race and the Fourth Amendment literature obscures not only the explicit ways in which law enforcement officials employ race (Latino identity) as a basis for suspicion (undocumentation), but also the role the Supreme Court has played in legitimizing that practice. Exacerbating immigration exceptionalism, the idea that immigration enforcement is not subject to ordinary constitutional constraints, the Court has given immigration officials (and indirectly law enforcement) wide discretion in enforcing immigration law in the interior. The general failure to engage the undocumented cases glosses over these dynamics and facilitates the claim that existing immigration enforcement practices do not rely on race, compounding the difficulties Latinos already experience framing their social vulnerabilities in racial terms. Finally, because criminal procedure scholars have relegated the undocumented cases to the borders of scholarship on race, racial profiling, and the Fourth Amendment, they have not explicated the ways in which the undocumented cases unduly expand police authority not only in the immigration enforcement context, but also in the investigatory domain more broadly.







- \* This article is an abridged version of Devon W. Carbado & Cheryl I. Harris, *Undocumented Criminal Procedure*, 58 UCLA L. REV. 1543 (2011).
- \*\* For comments on or conversations about this article, many thanks to Paul Butler, Kimberlé Crenshaw, Angela Davis, Ingrid Eagly, Laura Gómez, and Hiroshi Motomura. As always, the Hugh & Hazel Darling Law Library at UCLA School of Law provided valuable research assistance, and we especially want to thank Amy Atchison, Director of Reference and Research Services, for her guidance and support. Our research assistants, Adhana Davis, Antonio Kizzie, Will Pilon, Kim Sanders and Alicia Virani, also provided tremendous help. We thank Brittany Goodnight for helping us redact our original article for this venue.
1. See, e.g., Ingrid V. Eagly, *Prosecuting Immigration*, 104 NW. U. L. REV. 1281 (2010).
  2. See generally *id.*
  3. *Id.* at 1356.
  4. *Id.* at 1286.
  5. See Gabriel J. Chin, *Segregation's Last Stronghold: Race Discrimination and Constitutional Law of Immigration*, 46 UCLA L. REV. 1 (1998); Juan F. Perea, *Introduction to IMMIGRANTS OUT!: THE NEW NATIVISM AND THE ANTI-IMMIGRANT IMPULSE IN THE UNITED STATES* (Juan F. Perea ed., 1997).
  6. See also KEVIN R. JOHNSON, THE "HUDDLED MASSES" MYTH: IMMIGRATION AND CIVIL RIGHTS (2004); Jennifer Chacón, *Citizenship and Family: Revisiting Dred Scott*, 27 WASH. U. J.L. & POL'Y 45, 64 (2008).
  7. See Nicholas De Genova & Ana Y. Ramos-Zayas, *Latino Rehearsals: Racialization and the Politics of Citizenship Between Mexicans and Puerto Ricans in Chicago*, J. LATIN AM. ANTHROPOLOGY, June 2003, at 18, 21.
  8. See Neil Gotanda, *Comparative Racialization: Racial profiling and the Case of Wen Ho Lee*, 47 UCLA L. REV. 1689, 1691 (2000).
  9. We define racial profiling as "the practice of a law enforcement agent or agency relying, to any degree, on race . . . in selecting which individual to subject to routine or spontaneous investigatory activities." End Racial Profiling Act of 2007, S. 2481, 110th Cong.
  10. There are some notable exceptions to which we cite in the full article.
  11. See RANDALL KENNEDY, *Race, Law, and Suspicion: Using Race as a Proxy for Dangerousness*, in RACE, CRIME AND THE LAW 136-67, 159 (1997); see also JODY DAVID ARMOUR, NEGROPHOBIA AND REASONABLE RACISM: THE HIDDEN COSTS OF BEING BLACK IN AMERICA 13-14 (1997).
  12. See also David A. Harris, *The Stories, the Statistics, and the Law: Why "Driving While Black" Matters*, 84 MINN. L. REV. 265 (1999).
  13. 422 U.S. 873 (1975).
  14. 466 U.S. 210 (1984).
  15. 428 U.S. 543 (1976).
  16. We surveyed five leading criminal procedure casebooks to ascertain the extent to which the undocumented cases appear in them as compared with the documented cases.

17. 392 U.S. 1 (1968).
18. 517 U.S. 806 (1996).
19. 501 U.S. 429 (1991).
20. See Bernard Harcourt, *United States v. Brignoni-Ponce and United States v. Martinez-Fuerte: The Road to Racial Profiling*, in *CRIMINAL PROCEDURE STORIES* 315, 323-24 (Carol Steiker ed., 2006).
21. See Charles Becton, *The Drug Courier Profile: 'All Seems Infected That Th' Infected Spy, as All Looks Yellow to the Jaundic'd Eye*, 65 N.C. L. REV. 417, 426, 433-34 (1987).
22. See Harcourt, *supra* note 20, at 324.
23. *Id.* at 324-25.
24. *Id.*
25. *Id.*
26. Greg Williams, *Selective Targeting in Law Enforcement*, NAT'L B. ASS'N MAG., Mar./Apr. 1996, at 18.
27. We note here the limitations of this approach, as we have not disaggregated the data to determine whether an article that mentions racial profiling and Blacks also engages racial profiling and Latinos. Thus, the total frequency may not accurately reflect the actual number of articles. Nor have we engaged this analysis from a qualitative perspective—that is, whether the articles do more than mention Latino profiling as distinct from substantively engaging the issue.
28. Rebecca Porter, *Skin Deep: Minorities Seek Relief From Racial Profiling*, 35 TRIAL 13 (1999).
29. Numerous cases arising out of the experiences of Black motorists were brought charging various law enforcement agencies with racial profiling. See, e.g., Joint Application for Entry of Consent Decree, *United States v. New Jersey*, No. 99-5970 (D.N.J. Dec. 30, 1999).
30. Porter, *supra* note 28, at 15. JOSEPH NEVINS, OPERATION GATEKEEPER: THE RISE OF THE “ILLEGAL ALIEN” AND THE MAKING OF THE U.S.-MEXICO BOUNDARY III (2002).
31. See Muneer I. Ahmad, *A Rage Shared by All: Post-September 11 Racial Violence as Crimes of Passion*, 92 CALIF. L. REV. 1259 (2004); Susan M. Akram & Kevin R. Johnson, *Race, Civil Rights, and Immigration Law After September 11, 2001: The Targeting of Arabs and Muslims*, 58 N.Y.U. ANN. SURV. AM. L. 295 (2002); Asli Ü. Bâli, *Changes in Immigration Law and Practice After September 11: A Practitioner's Perspective*, 2 CARDOZO PUB. L. POL'Y & ETHICS J. 161 (2003); Bernard E. Harcourt, *Muslim Profiles Post 9/11: Is Racial Profiling an Effective Counterterrorist Measure and Does It Violate the Right to Be Free From Discrimination?* (Univ. of Chi. Law Sch. John M. Olin Law & Economics Working Paper No. 288, 2006), available at <http://www.law.uchicago.edu/files/files/286.pdf><http://www.law.uchicago.edu/files/files/286.pdf>.
32. DAVID A. HARRIS, PROFILES IN INJUSTICE: WHY RACIAL PROFILING CANNOT WORK (2002); Harcourt, *supra* note 20. David Cole's earlier book also noted that Latinos near the border were heavily burdened by racially targeted enforcement. See DAVID COLE, NO EQUAL JUSTICE: RACE AND CLASS IN THE AMERICAN CRIMINAL JUSTICE SYSTEM (1999).
33. See, e.g., David A. Harris, *Using Race or Ethnicity as a Factor in Assessing the Reasonableness*

- of Fourth Amendment Activity: Description, Yes; Prediction, No*, 73 Miss. L.J. 423 (2003).
34. This section draws on the work of LAURA GÓMEZ, *MANIFEST DESTINIES: THE MAKING OF THE MEXICAN AMERICAN RACE* (2007).
  35. See GEORGE J. SANCHEZ, *BECOMING MEXICAN AMERICAN: ETHNICITY, CULTURE, AND IDENTITY IN CHICANO LOS ANGELES, 1900–1945*, at 38–62 (1995).
  36. See GÓMEZ, *supra* note 34, at 3–4.
  37. Shortly after ratification of the Constitution, the first Congress adopted the Naturalization Act of 1790, restricting naturalization to “free white persons.” This restriction remained in effect until 1870 when the restriction was amended to permit “aliens of African nativity and . . . persons of African descent.” HIROSHI MOTOMURA, *AMERICANS IN WAITING: THE LOST STORY OF IMMIGRATION AND CITIZENSHIP IN THE UNITED STATES 173* (2006). Racial restrictions on other nonwhites remained until 1952. *Id.* at 75.
  38. GÓMEZ, *supra* note 34, at 17–18, 41–45.
  39. *Id.* at 1, 83–85.
  40. As GÓMEZ illustrated, this perception of Mexican Americans as “off-white” or less than fully white obtained even in New Mexico where Mexican men enjoyed a certain range of political rights (jury service, voting, and holding office in the territorial legislature). *Id.* at 83–90. See also IAN HANEY LÓPEZ, *RACISM ON TRIAL: THE CHICANO FIGHT FOR JUSTICE* (2003).
  41. GÓMEZ, *supra* note 34, at 142–43.
  42. 81 F. 337 (W.D. Tex. 1897).
  43. GÓMEZ, *supra* note 34, at 140.
  44. *Id.* at 139–41.
  45. *Id.* at 140–41.
  46. *Id.* at 141.
  47. See Neil Foley, *Over the Rainbow: Hernandez v. Texas, Brown v. Board of Education, and Black v. Brown*, 25 CHICANO–LATINO L. REV. 139, 140–41 (2005).
  48. 161 F.2d 774 (9th Cir. 1947).
  49. *Id.* at 780 (emphasis added).
  50. 347 U.S. 475 (1954).
  51. 347 U.S. 483 (1954).
  52. *Hernandez*, 347 U.S. at 480.
  53. This argument was accepted by the Texas appellate court, which had rejected the assertion that Hernandez’s rights had been violated: “Mexicans are white people . . . it cannot be said in the absence of proof of actual discrimination that appellant had been discriminated against.” *Hernandez v. State*, 251 S.W.2d 531, 536 (Tex. Crim. App. 1952), *rev’d*, *Hernandez*, 347 U.S. 475.
  54. *Hernandez*, 347 U.S. at 479.
  55. 198 P.2d 17 (Cal. 1948).
  56. *Id.* at 29.
  57. 388 U.S. 1 (1967).
  58. *United States v. Brignoni-Ponce*, 422 U.S. 873, 887 (1975).

59. Indeed, the asserted emergence of a consensus against racial profiling prior to the 9/11 attacks largely rested upon differing conceptions about what racial profiling entailed.
60. See Harcourt, *supra* note 20, at 317 n.6, 317–18. Some jurisdictions define the term narrowly to include investigatory practices based solely on race, while others prohibit any consideration of race. Compare MD. CODE ANN., TRANSP. § 25-113 (Michie Supp. 2001), and R.I. GEN. LAWS § 31-21.1-2 (2000), with Consent Decree, *Wilkins v. Md. State Police*, No. CCB-93-468 § 2.1 (D. Md. 1993) (consent decree prohibiting racial profiling in case brought by the NAACP and the ACLU against the Maryland State Police), and *Constitutional Issues*, TUCSON POLICE DEP'T (revised July 29, 2010), [http://tpdinternet.tucsonaz.gov/general\\_orders/2200CONSTITUTIONAL%20ISSUES.pdf](http://tpdinternet.tucsonaz.gov/general_orders/2200CONSTITUTIONAL%20ISSUES.pdf). For consideration of this debate over meaning, see Kathryn K. Russell, *Racial Profiling: A Status Report of the Legal, Legislative, and Empirical Literature*, 3 RUTGERS RACE & L. REV. 61, 65–68 (2001).
61. We adopt the definition that encompasses all uses of race, as the utilization of other factors does not erase the fact that a racial profile—that association between racial identity and the suspicion of some illegal behavior—is still present.
62. See De Genova & Ramos-Zayas, *supra* note 7.
63. See Ruben J. Garcia, *Critical Race Theory and Proposition 187: The Racial Politics of Immigration Law*, 17 CHICANO-LATINO L. REV. 118, 119 (1995); Kevin R. Johnson, *How Racial Profiling in America Became the Law of the Land: United States v. Brignoni-Ponce and Whren v. United States and the Need for Truly Rebellious Lawyering*, 98 GEO. L.J. 1005, 1038 & n.196 (2010). In actuality, most Latinos are citizens or legal resident aliens. See *The American Community—Hispanic: 2004*, U.S. CENSUS BUREAU, at 11 (2007), available at <http://www.census.gov/prod/2007pubs/acs-03.pdf>.
64. KELLY LYTLE HERNÁNDEZ, *MIGRA! A HISTORY OF THE U.S. BORDER PATROL* 9 (2010); see also *Border Patrol History*, CBP.GOV, [http://www.cbp.gov/xp/cgov/border\\_security/border\\_patrol/border\\_patrol\\_ohs/history.xml](http://www.cbp.gov/xp/cgov/border_security/border_patrol/border_patrol_ohs/history.xml) (last visited July 12, 2011).
65. LYTLE HERNÁNDEZ, *supra* note 64, at 32–36.
66. See Kristin Connor, *Updating Brignoni-Ponce: A Critical Analysis of Race-Based Immigration Enforcement*, 11 N.Y.U. J. LEGIS. & PUB. POL'Y 567, 586 (2008). 45 percent of undocumented immigrants are already in the United States when they go into undocumented status, mostly by overstaying their visas. The majority of this population is non-Latino. *Id.* at 587.
67. LYTLE HERNÁNDEZ, *supra* note 64, at 44.
68. *Id.* at 41–42.
69. As Lytle Hernández puts it:  
 The Border Patrol's narrow focus upon policing unsanctioned Mexican immigration . . . drew a very particular color line around the political condition of illegality. Border Patrol practice, in other words, imported the borderlands' deeply rooted racial divides arising from conquest and capitalist economic development into the making of U.S. immigration law enforcement and, in turn, transformed the legal/

illegal divide into a problem of race.

*Id.* at 222.

70. *Id.* at 10.

71. *Id.* at 9.

72. The Border Patrol offered statistics in *Brignoni-Ponce* asserting that 85 percent of the persons arrested for illegal entry were people of Mexican origin. *United States v. Brignoni-Ponce*, 422 U.S. 873, 879 (1975).

73. JOHNSON, *supra* note 6, at 48–49.

74. *See* Chin, *supra* note 5, at 5.

75. *See id.*; *see also* IAN HANEY LÓPEZ, *WHITE BY LAW: THE LEGAL CONSTRUCTION OF RACE* (10th ed. 2006); Devon W. Carbado, *Yellow by Law*, 97 CALIF. L. REV. 633 (2009).

76. *See* Kerry Abrams, *Polygamy, Prostitution, and the Federalization of Immigration Law*, 105 COLUM. L. REV. 641 (2004); John Hayakawa Torok, *Reconstruction and Racial Nativism: Chinese Immigrants and the Debates on the Thirteenth, Fourteenth, and Fifteenth Amendments and Civil Rights Laws*, 3 ASIAN L.J. 55, 96–97 (1996).

77. Torok, *supra* note 76, at 96.

78. *See* Gabriel Chin, *The Civil Rights Revolution Comes to Immigration Law: A New Look at the Immigration and Nationality Act of 1965*, 75 N.C. L. REV. 273, 279 (1996). For a more detailed description of the national origins system, *see* MOTOMURA, *supra* note 37, at 126–30.

79. 130 U.S. 581 (1889).

80. *Id.* at 609.

81. *Id.* at 595.

82. 149 U.S. 698 (1893).

83. *See* Chin, *supra* note 5, at 18.

84. *See* JOHNSON, *supra* note 6, at 46.

85. *See* *Zadvydas v. Davis*, 533 U.S. 678, 678, 695 (2001); MOTOMURA, *supra* note 37, at 102–08.

86. *See* Chin, *supra* note 5, at 3–4, 9; Janel Thamkul, *The Plenary Power–Shaped Hole in the Core Constitutional Law Curriculum: Exclusion, Unequal Protection, and American National Identity*, 96 CALIF. L. REV. 553, 558, 575–78 (2008).





Laura E. Gómez

Professor of Law

Laura Gómez teaches in the areas of race and the law, law and society, constitutional law, civil procedure and criminal law. Her scholarship, including two books, has focused on the intersection of law, politics and social stratification in both contemporary and historical contexts. In *Misconceiving Mothers: Legislators, Prosecutors and the Politics of Prenatal Drug Exposure* (1997), she documented the career of the “crack baby” / “crack mother” social problem in the media and public policy, situating it at the nexus of the abortion debate, the drug war and competing discourses of criminalization and medicalization as they played out in the late 1980s. In her 2007 book, *Manifest Destinies: The Making of the Mexican American Race*, Gómez examines how law and racial ideology intersected to create new racial groups and to restructure the turn-of-the-twentieth-century racial order in the Southwest and the nation as a whole.

She is active in national scholarly organizations, including the Law and Society Association, where she just completed a two-year term as president. She has served as an associate editor of *Law & Society Review* and has served on the editorial boards of *Law & Social Inquiry*, *SIGNS*, *Aztlan* and *Studies in Law, Politics and Society*.

Professor Gómez rejoined the UCLA faculty in 2011, following six years as a professor of law and American studies at the University of New Mexico. From 1994 to 2005, she held joint appointments in law and sociology at UCLA. From 2000 to 2002, she was a co-founder and the first co-director of UCLA’s Critical Race Studies Specialization. In addition to master’s and doctoral degrees in sociology from Stanford, she earned a J.D. from Stanford Law School. She clerked on the Ninth Circuit Court of Appeals for Judge Dorothy W. Nelson.





# UNDERSTANDING LAW AND RACE AS MUTUALLY CONSTITUTIVE\*

Laura E. Gómez

This article examines the relationship between law and race, highlighting the fact that law and race shape each other in powerful ways that until recently have been little explored by scholars. Social scientists who study law have tended to focus on race as an independent variable that helps predict a legal outcome and have often narrowly defined race as phenotype and measured it in binary (Black or White) terms.<sup>1</sup> Critical race theorists (mostly legal scholars), in contrast, have made race their central focus and have treated law as an independent variable that explains race, in its various manifestations, though they have not tended to systematically use social science methods.<sup>2</sup>

## I. INTRODUCTION

Critical race theory emerged in the mid-1980s, along with a critical mass of African American law professors in the U.S.<sup>3</sup> Scholars in the field write in the areas of civil rights and race law (with particular prominence in the fields of constitutional law, anti-discrimination law, and employment law), as well as in an increasingly diverse array of other doctrinal areas such as criminal law and procedure, torts, family law, tax law, and environmental law.<sup>4</sup> Critical race scholars write about race and the law from a perspective that is critical of the anti-discrimination model that has been dominant in legal scholarship and American jurisprudence since the 1970s. The anti-discrimination model conceives of racism and racial discrimination as individualized, aberrational, and capable of remedy within the current legal framework, whereas critical race scholars view racism as institutionalized and endemic, and thus as frequently immune to anti-discrimination law and policy.

The past two decades have seen the rise of a literature that looks deeply at the role played by law in constructing racial identities and categories, as well as compared how law has shaped the experiences of different racial groups in the U.S.<sup>5</sup> A related recent literature has explored law's role in shaping ostensibly non-racial categories that are heavily endowed with considerable racial meaning, such as citizens, criminals, drug addicts, terrorists, etc.<sup>6</sup>

More recently, scholars have recognized that the constitutive process goes in both directions: "law not only constructs race, but race constructs law: racial conflicts dis-

tort the drafting and implementation of laws; skew the development, character and mission of legal bureaucracies; alter how various communities, including Whites, understand and interact with legal institutions; and twist the self-conception of legal actors, from lawmakers to lawyers, cops to judges.”<sup>7</sup> This article identifies an emerging genre of socio-legal scholarship that explores how law and race construct each other in an ongoing, dialectic process that ultimately reproduces and transforms racial inequality.

I draw on several recently published monographs that should be seen as constituting this emerging literature. Each book focuses on the interaction of law and race in a particular time and place (most in the U.S., including regions colonized by the U.S. such as Hawaii and the Southwest, but also in Canada and Jamaica). Comparing in-depth case studies of how law and race intersect in particular geographic locations at particular times allows us to begin to tease out some general patterns that describe how law and race mutually constitute each other. Though these are historical studies, they are written by scholars trained in the disciplines of anthropology, law, political science, sociology, and history (and combinations of those fields), as well as in interdisciplinary doctoral programs.

Many books about race take up the law to some extent, but I have included only studies that both feature law in a central way and treat law as a dynamic social and cultural force. *Law* is broadly defined to include legislation, appellate opinions, trials, litigation/prosecution data, and the activities of legal actors (both formal, such as judges and lawyers, and informal, such as mid- and low-level bureaucrats who initiate, implement or support legal processes). I have not included books concerned primarily with legal doctrine or its evolution.<sup>8</sup> A second selection criterion was how the books approach race: these books embrace the intellectual study of race as a socially constructed phenomenon, and they appreciate law as a central force shaping race. *Race* is understood broadly to refer to a range of social phenomena that can be operationalized as racial categories and boundaries, racial identity (including how race intersects with gender, class, and sexual and other identities), racial conflict, racial ideology, racism, and so forth.<sup>9</sup>

While I focus here on historical case studies, I am not arguing that law and race are mutually constitutive only in the past. Law and race continue to interact in powerful ways today,<sup>10</sup> but there is something compelling about historical examples of their interaction. In part this is because examples from history allow us to unmask race as a part of the natural world we take for granted; they invite us to step out of our own social world where, in general, it is harder for the beneficiaries of White privilege to see race and racism being enacted. Historical cases are also appealing because there is little contention over the general fact that law played a central role

in producing and reproducing racial subordination in the past, while there is much more debate over law's current role in reproducing racism.<sup>11</sup>

There also are unique methodological advantages to historical research: interpretation of archival materials may allow social scientists a great deal more latitude in how they capture race, by measuring it variously as racial categories and boundaries, racial identity, racial conflict, racial ideology, and racism, rather than as a dichotomous variable based on self-identification, for example. Moreover, historical methods have been embraced by legal scholars in the critical race theory movement, even when those scholars have not had formal training in history. But an emphasis on historical cases also introduces particular hazards, the most significant of which is the danger of presentism—of blithely applying contemporary ideas about race to historical contexts where they simply were not relevant. I have tried to guard against the latter and to point out where I think authors have made that error.

**W**hat race means is deeply contested in popular culture, law, politics, and science. Sociologist Ann Morning's research on contemporary popular conceptions of race finds three dominant understandings of race: (1) race as biology (despite the fact that scientists agree that race is not biologically meaningful, people continue to believe, according to Morning's research, that biology produces "real" racial differences); (2) race as culture (people associate racial difference with cultural differences like musical preferences, food preferences, celebration of cultural traditions; here race is used similarly to ethnicity); and (3) socially constructed race (the idea that race is not real but rather produced by people; this can support a conservative orientation like color blindness or a progressive orientation like affirmative action).<sup>12</sup> Sociologist Eduardo Bonilla-Silva identified a fourth popular notion of race: color blind race. Under this view, the social constructionist view of race combines with the notion that people should pretend they do not "see" race and/or, if they do see race, ignore it.<sup>13</sup>

## II. THEORIZING RACE

These four popular conceptions of race also infuse scientific, political, and legal notions of race.<sup>14</sup> For example, the Supreme Court in its case law on the 14th Amendment and anti-discrimination legislation like Title VII shifts back and forth between a variety of the notions of race.<sup>15</sup> Moreover, these very different ideas about race are not mutually exclusive for most people, but instead coexist, only to be situationally invoked by people to make sense of everyday interactions in which race is salient.<sup>16</sup>

The concept of race "invokes biologically based human characteristics (so-called 'phenotypes'), [but] selection of these particular human features for racial signifi-

tion is always and necessarily a social and historical process.<sup>17</sup> This view of race as socially constructed emphasizes power relations (subordination) and inequality (stratification), drawing heavily on the historical roots of racial exclusion, rather than, for example, racial identity. From this view, the historical facts of racial exclusion are paramount—exclusion from personhood under slavery; exclusion from citizenship in cases such as *Dred Scot*, in laws that restricted naturalization to Whites (and after the Civil War, to Whites and Blacks), and in the contemporary demonization of Mexican immigrants as undeserving of citizenship; exclusion from particular spaces via Jim Crow legislation; and exclusion from political rights such as voting, serving on juries, running for elected office, testifying in court, and so forth. “In that history, racial classification turned not on what one felt [or how one identified racially], but, instead, on what others allowed one to do.”<sup>18</sup>

While both race and ethnicity are about socially constructed group difference in society, race is always about hierarchical social difference, whereas ethnicity may be non-hierarchical, depending on the social context. While it has become common for social scientists to prefer to talk about race as ethnicity since the late 1940s,<sup>19</sup> I eschew the term ethnicity because its use tends to defuse the emphasis on race as fundamentally about power and stratification.<sup>20</sup> In particular, a focus on ethnicity tends to emphasize individualized self-identification as an unfettered choice, rather than the structural constraints on racial boundaries that exist as the result of historically-rooted racial oppression. Despite the fact that ethnicity and race overlap considerably and the ways in which ethnicity remains important in diverse societies across the world, I agree with sociologists Stephen Cornell and Douglas Hartmann that race remains “the most powerful and persistent group boundary in American history.”<sup>21</sup>

The studies highlighted here adopt the social constructionist view of race that has become pervasive in the social sciences over the past few decades.<sup>22</sup> As an intellectual approach to race, the constructionist view has three main components. First, it rejects a biological basis for race (i.e., “there is greater variation within racial groups than between them”).<sup>23</sup> Second, it views race as a social construct: “a social invention that changes as political, economic, and historical contexts change.”<sup>24</sup> Third, although race is socially constructed (indeed, because of its power as a social construct), race has real consequences.<sup>25</sup> In its recent statement on the topic, the American Sociological Association concluded that race is embedded in virtually all American social institutions and practices.<sup>26</sup>

One of the most compelling theories taking the constructionist position is the theory of racial formation put forth in 1986 by sociologists Michael Omi and Howard Winant.<sup>27</sup> According to this theory, race has both ideological and structural dimen-

sions: “A vast web of racial projects mediates between the discursive or representational means in which race is identified and signified on the one hand, and the institutional and organizational forms in which it is routinized and standardized on the other.”<sup>28</sup> The state (state institutions, state actors, government agencies and policies) plays a major role in structuring race and racism and, as a result, law is a key player in racial formation theory, despite the fact that Omi and Winant did not develop that line of analysis.

The studies in this emerging field build, self-consciously or not, on racial formation theory by situating racial projects within legal systems and processes. In this way, they contribute to our knowledge about how law and race both relate to broader political dynamics and state projects. I proceed by describing how these studies illustrate the process by which law and race co-construct each other in a continuous, back and forth process. Cumulatively, they present new insights about law, race, and how they co-construct each other to ultimately reproduce and transform racial inequality in society.

Sociologist Renisa Mawani’s recent book *Colonial Proximities: Crossracial Encounters and Juridical Truths in British Columbia, 1871-1921*<sup>29</sup> provides a gripping portrait of how the racial order and the legal order shaped each other in 19th century British Columbia. She describes this “colonial contact zone” as “a space of racial inter-mixture—a place where Europeans, aboriginal peoples, and racial migrants came into frequent contact, a conceptual and material geography where racial categories and racisms were both produced and productive of locally configured and globally inflected modalities of colonial power” and where government officials, missionaries, and private employers (who exercised a quasi-legal authority) generated practices of colonial governance that were fundamentally racialized.<sup>30</sup> Two examples illustrate how racial dynamics shaped law and how law in turn shaped the racial order.

### III. LAW AND THE RACIAL ORDER

Some of the earliest sites for inter-racial encounters were the salmon canneries, central to the region’s capitalist development and hence its emergence as a viable colonial outpost. The canneries relied on a racially diverse workforce that included mostly male White settlers, male Chinese immigrants, and local aboriginal people (both men and women). Inter-racial mixture at work threatened White domination by producing mixed race progeny and by introducing the potential for inter-racial solidarity among workers.<sup>31</sup> One of the mechanisms employed by the cannery owners to decrease these possibilities—in a setting in which they regulated many aspects of workers’ lives à la company towns—was to assign housing by race.<sup>32</sup> Among Whites, housing was segregated by class as well, with White elites living in

larger, single-family homes (located furthest from the worst odors of the cannery) and White workers assigned to private bungalows and cottages. Chinese workers, in contrast, were forced to reside collectively in “overcrowded and unsanitary” bunkhouses; and aboriginal workers were pushed either to the outskirts of the canneries where they worked or remained living in their nearby villages, often located on the periphery of canneries. In this way, law-like residential segregation inscribed pre-existing racial differences in order to sharpen those differences (and dampen cross-racial contact) in a newly racially diverse geographic setting.

Another example comes from the legal regulation of prostitution. Putatively, prostitution anxiety focused on aboriginal, mixed race, and Chinese girls and women who were perceived as being exploited by aboriginal and Chinese men who sold “their women” into the sex trade (or allowed women to sell themselves into it). Contrasting and dynamic state responses to aboriginal and Chinese women illustrate how law and the racial order produced each other. In the early contact period characterized by European fur traders, there was no effort to regulate White men’s sexual and social relations with aboriginal women.<sup>33</sup> Later, when the numbers of White female settlers increased significantly, colonial authorities promoted inter-racial prostitution rather than concubinage in order to encourage White endogamy.<sup>34</sup> By the end of the century, after an express legal campaign, aboriginal women were contained on reserves and were no longer perceived as a marital or sexual threat to settler society.<sup>35</sup> But by that time, the newer population of Chinese immigrants was perceived as “contaminating” settler society.<sup>36</sup> Based on her review of the correspondence, legislation, and other official documents written to and by colonial officials, Mawani concludes that the late 19th and early 20th century anti-prostitution rhetoric became the justification for the physical exclusion of Chinese immigrants at the border (especially female immigrants), as well as a way to justify the continued political exclusion of those Chinese who already had entered British Columbia.<sup>37</sup> Thus, legal responses to prostitution themselves hardened racist ideas, while simultaneously reflecting taken-for-granted racial truths.

My book *Manifest Destinies: The Making of the Mexican American Race*<sup>38</sup> explores a different colonial contact zone, 19th century New Mexico, but similarly looks at how law and race interacted and ultimately reproduced and transformed racial inequality. In a setting in which American colonizers had neither a realistic chance of militarily dominating large numbers of native Mexicans and diverse Indian peoples or the hope of quickly attracting large numbers of White settlers, they embraced a divide and conquer strategy in which whiteness became a key wedge between Mexicans and Pueblo Indians.

Building on the pre-existing Spanish-Mexican racial order, the Americans exploited Mexicans’ claims of racial mixture (as a people descended both from Spaniards and

Indians) to justify endowing Mexican men with a host of rights (voting rights, the right to hold office, jury service, etc.) and to withhold these same rights from Pueblo Indian men, even though the latter had citizenship rights under Mexican rule and arguably under the treaty ending the 1846-'48 war between the U.S. and Mexico.<sup>39</sup> The result was a local racial order in which Mexican Americans functioned as a wedge group between White Americans, located above them on the racial hierarchy, and Pueblo Indians, located below them. At the national level, Mexican Americans again played a wedge role due to their off-white status, buffering Whites above them (and especially marginal Whites like Irish and Italian immigrants) from Blacks at the bottom of the racial order.<sup>40</sup>

In his book *Racism on Trial: The Chicano Fight for Justice*,<sup>41</sup> legal scholar Ian Haney López explores the 20th century ramifications of Mexican Americans' 19th century status as an off-white racial group. Although others believe he overstates the case,<sup>42</sup> Haney López argues that Mexican Americans were poised at the time of the Chicano civil rights movement to choose between a White and a non-white racial identity. The larger society's view of Mexican Americans as non-white others played a crucial role, especially as it was manifested in responses by police and prosecutors in two criminal trials of groups of young Mexican American men for politically motivated offenses in the early 1970s.

Haney López postulates a theory about how racial ideology is reproduced as a key aspect of producing the racial order: "[H]ow do ideas about race operate—how do they arise, spread, and gain acceptance? What is the relationship between race as a set of ideas and racism as a set of practices?"<sup>43</sup> Building on Omi and Winant's work, he postulates that "common sense racism"—"a complex set of background ideas that people draw on but rarely question in their daily affairs ... stock ideas and practices that we have absorbed and heavily relied upon but to which we give little thought"—provides the answer.<sup>44</sup> For example, the taken-for-granted notion that Mexican Americans were generally inferior to Whites (common sense in mid-20th century California) led Los Angeles County judges to exclude them from grand jury service, even as they proclaimed that they did not personally know any qualified Mexican Americans and that they did not intend to discriminate against Mexican Americans.<sup>45</sup> The racial order virtually ensured the legal system's exclusion of Mexican American citizens on grand and petit juries, and that legal outcome, in turn, affirmed their racially inferior position in society.

Anthropologist Pem Davidson Buck similarly explores the ways in which race becomes naturalized after decades of the common-sense reproduction of racist ideas, in this case ideas about race deeply intertwined with class-based stereotypes.



In *Worked to the Bone: Race, Class, and Privilege in Kentucky*,<sup>46</sup> Buck notes that Kentucky's early homesteads went only to White veterans, but with a built-in bias based on social class: enlisted men received 100 – 300 acre lots, whereas officers sometimes received thousands of acres.<sup>47</sup> During this era, homesteading was risky because the region's original inhabitants, Cherokee and Shawnee Indians, adamantly resisted White encroachment on their lands [they continued to do so until they were forcibly removed to Indian Territory (later Oklahoma) in the 1830s]. According to Buck, poor White settlers in Kentucky constructed themselves racially against these Indian populations: "In essence they became a military buffer between Native Americans and advancing White settlement. For people without access to capital or land in the heavily settled East, the chance for upward mobility—if they survived—made the risk worthwhile. They now had reason to treasure White privilege."<sup>48</sup>

But the precariousness of frontier life, coupled with rampant land speculation, meant that property quickly became concentrated among wealthy Whites: by 1780, 75 percent of White Kentuckians were poor and landless,<sup>49</sup> and within another two decades, 21 White landowners owned one-quarter of the state's land.<sup>50</sup> Some delegates to the constitutional convention of 1792 argued that the franchise should be restricted to property owners, but given the land distribution, that probably would have led to a revolt among the White masses. Instead, in a move that would have repercussions for the next two centuries, all White men were enfranchised as a way to solidify White privilege and the Black/White racial divide.<sup>51</sup>

Political scientist Julie Novkov explores similar themes in a very different style in her 2008 book *Racial Union: Law, Intimacy, and the White State in Alabama, 1865-1954*.<sup>52</sup> She rejects ahistorical invocations of "white supremacy," instead seeking to link racial ideology to state-building in order "to describe the linkage between racial ideology in politics and culture and its concrete manifestations in state institutions in the post-bellum U.S. South."<sup>53</sup> She examines anti-miscegenation law and its enforcement as a key site "for the creation, articulation, rationalization, and ultimately reflection of the supremacist state, through its attention to the meaning of racial boundaries."<sup>54</sup> Novkov persuasively illustrates how racial ideology (White supremacy) produced racist laws (inter-marriage bans), and how that subsequently led to hardened racial boundaries that ultimately justified the racial order in which Blacks were subordinate to Whites.

Delegates to the 1901 state constitutional convention vigorously debated but ultimately rejected two amendments to the anti-miscegenation law: one that would have defined Blacks via the hypodescent rule and a second that would have added Chinese and Native Americans to those proscribed from marrying Whites. Instead, they effectuated the subordination of Blacks by voting to disenfranchise African

American men; within two years, the number of Black men registered to vote plummeted from 181,000 to 5,000.<sup>55</sup> The first quarter of the century witnessed a series of anti-miscegenation cases (trial and appellate) in which defendants raised a variety of definitional challenges to their status as “White” or “Black” and, faced with inconsistent responses from the courts, the Alabama legislature in 1927 adopted the one-drop rule—any Black ancestry sufficed to make a person Black.<sup>56</sup> Law both reflected the racial order and helped to produce it in a more intransigent form.

While legal narratives of the American South often focus exclusively on Black/White race relations, both Buck and Novkov are attentive to the presence of American Indians in the South and to the attendant complications of a multi-group racial order. Historian Moon-Ho Jung more directly takes up questions of a tri-racial dynamic in the U.S. South by interrogating the ideological and material roles of “coolies”—exploited Chinese contract laborers—in post-bellum Louisiana. In *Coolies and Cane: Race, Labor and Sugar in the Age of Emancipation*,<sup>57</sup> Jung links 19th century immigration law and policy with the national dialogue about slavery and emancipation, while also putting the South in the broader context of both Caribbean sugar production (Louisiana’s main competitor at the time) and Chinese migration across the Western hemisphere.

Sugar plantation owners needed a large, flexible workforce [many more workers were needed during the grinding and planting seasons than at other times<sup>58</sup>], so Louisiana planters turned to Chinese laborers as a way to provide flexibility after emancipation so that they would not be dependent on recently freed slaves. Jung investigates one sugar plantation’s labor policies immediately following emancipation, finding that their hiring rolls included free Blacks (who worked at wages ranging from \$8 – \$19.50/month), White (European) immigrants (who contracted for \$20/month pay and their transportation costs from Chicago, if they stayed four months or more), and Chinese contract laborers (who worked for \$16/month).<sup>59</sup> Ironically, Louisiana planters’ labor shortage became even more acute after 1877, when Republican rule was defeated in Louisiana and a Black exodus to Kansas led planters to depend even more on immigrant laborers, both White and Chinese.<sup>60</sup> In the end, coolies served as a surplus army of labor for sugar planters in Louisiana, even as they played a role in “whitening” otherwise marginal European immigrants who moved to the region in the post-bellum period.<sup>61</sup>

Anthropologist Virginia Domínguez presents a fascinating study of the complex ways in which individual identity choices are heavily constrained by both social meaning and institutional forces (including the legal system) in *White by Definition: Social Classification in Creole Louisiana*.<sup>62</sup> A system of racial hierarchy that accreted over centuries (and three different colonial governments) eventually was codified

via Louisiana's anti-miscegenation laws, themselves designed to limit the inter-generational transfer of wealth from White men to women of color (and their mixed-race children).<sup>63</sup> Statutory law and case law interacted in sometimes unpredictable (or perhaps highly predictable) ways. In 1910 the state supreme court ruled that a White man had not violated the law by living with a woman who was one-eighth Black because an "octoroon was not negro," but rather a person of color within the Louisiana tradition. Within 30 days of the ruling, the legislature banned unions between Whites and anyone who had any amount of African ancestry, thereby helping to solidify a new understanding of race as binary rather than tertiary.<sup>64</sup>

As the one-drop rule became entrenched in Louisiana, legal bureaucrats saw it as their obligation to enforce it. Domínguez tells of Louisiana's vital statistics registrar, Naomi Drake, who in the 1950s and 1960s instituted what she termed "race-flagging" of birth and death certificates. Drake investigated as racially "suspicious" 4,700 birth certificates and 1,100 death certificates between 1960 and 1965 alone.<sup>65</sup> Her enforcement did not stop with the passage of the federal Civil Rights Act or with the social changes in race relations of the 1970s, but only in 1983, when the state legislature mandated self-identification as Louisiana's definitive method of assignment to racial categories.<sup>66</sup> As historian Peggy Pascoe notes, bureaucrats like Drake played as significant a role as other legal actors (legislators, judges and prosecutors) in reproducing race and racism: "[officials like marriage license clerks] carried out their tasks as a matter of bureaucratic routine rather than criminal enforcement, in quiet county offices rather than dramatic courtrooms ... a seemingly natural documentary 'fact' of race was produced in marriage license bureaus."<sup>67</sup>

J. Kehaulani Kauanui, who has a doctorate from the History of Consciousness program at the University of California at Santa Cruz and who teaches in an anthropology department, has recently published a study of the congressional passage of the 1921 Hawaiian Homes Commission Act. Congress enacted the legislation roughly mid-way between the United States' formal acquisition of Hawaii as a colony in 1898—although American missionaries and business interests had been active on the islands since the 1820s<sup>68</sup>—and admission of Hawaii as a state in 1959. In *Hawaiian Blood: Colonialism and the Politics of Sovereignty and Indigeneity*,<sup>69</sup> Kauanui argues persuasively that Hawaii's racial order shaped the law's definition of who was Native Hawaiian for purposes of receiving land allotments under the law, which in turn came to define (and still often defines today) the category of Native Hawaiians under a 50 percent blood quantum rule. Hawaii's three-tiered racial order in the early 20th century consisted of Asians (who were typed as alien, non-citizens), Native Hawaiians, and everyone else, principally Whites but also mixed-race persons who did not fit squarely in the other categories.<sup>70</sup>

The Homes Commission law was ostensibly designed to provide redress to Native Hawaiians, who were suffering from drastic poverty and high mortality rates<sup>71</sup> and who were viewed collectively as capable of eventually assimilating into the White settler society.<sup>72</sup> In contrast, Asians (and especially the Japanese, who were numerically dominant at the time), were viewed as collectively unassimilable and also disenfranchised because Congress had refused to extend citizenship rights to Asian immigrants living in Hawaii in 1900 when it granted them to Native Hawaiians (U.S.-born Japanese Americans did not vote in substantial numbers until the 1930s).<sup>73</sup> The law's focus on providing reparations to Native Hawaiians (via land allotments) sought to rehabilitate them as against Asians, but sought to do so narrowly, setting a 50 percent blood quantum definition for Native Hawaiian status. Hawaii's racial dynamics produced the law, but the law exerted a powerful influence on those very dynamics by instituting a rigid definition of ancestry to define Native status (and by rejecting indigenous ideas about kinship that Kauanui argues today trump blood quantum in some contexts).<sup>74</sup> The result was the transfer of property wealth to a narrower segment of those who could potentially claim Native Hawaiian status, which left much of the land originally allotted in the public domain and thus available to be leased by sugar plantation owners and eventually owned Whites.<sup>75</sup>

**T**his article has described an emerging field in socio-legal studies that investigates how law and race mutually constitute each other in an ongoing, dialectic process. These studies build on racial formation theory by situating racial projects within legal systems and processes. In this way, they contribute to our knowledge about how law and race both relate to broader political dynamics and state projects. They illustrate how law and race co-construct each other in a continuous, back and forth process. Cumulatively, they present new insights about law, race, and how they co-construct each other to ultimately reproduce and transform racial inequality in society.

#### IV. CONCLUSION



- \* This essay is an abridged version of Laura E. Gómez, *Understanding Law and Race as Mutually Constitutive: An Invitation to Explore an Emerging Field*, 6 ANN. REV. L. & SOC. SCI. 487 (2010).
1. See Laura Gómez, *A Tale of Two Genres: On the Real and Ideal Links Between Law and Society and Critical Race Theory*, in THE BLACKWELL COMPANION TO LAW AND SOCIETY 453 (Austin Sarat ed., 2004); Osagie K. Obasogie, *Race in Law and Society: A Critique*, in RACE, LAW AND SOCIETY 445 (Ian Haney López ed., 2007). I follow historian Peggy Pascoe's lead in capitalizing Black(s) and White(s) because other terms frequently used as racial terms (American Indian(s), Mexican American(s), etc.) are capitalized and, for White(s), capitalization "help[s] mark the category that so often remains unmarked, and taken for the norm." PEGGY PASCOE, WHAT COMES NATURALLY: MISCEGENATION LAW AND THE MAKING OF RACE IN AMERICA 14 (2009).
  2. See Gómez, *supra* note 1.
  3. For edited collections on critical race theory, see DERRICK A. BELL, RACE, RACISM AND AMERICAN LAW (6th ed. 2008); DOROTHY A. BROWN, CRITICAL RACE THEORY: CASES, MATERIALS AND PROBLEMS (2003); CRITICAL RACE FEMINISM: A READER (Adrien Katherine Wing ed., 2d ed. 2003); CRITICAL RACE THEORY: THE CUTTING EDGE (Richard Delgado ed., 1995); CRITICAL RACE THEORY: THE KEY WRITINGS THAT FORMED THE MOVEMENT (Kimberlé Crenshaw et al. eds., 1995); CROSSROADS, DIRECTIONS AND A NEW CRITICAL RACE THEORY (Francisco Valdes et al. eds., 2002); JUAN F. PEREA ET AL., RACE AND RACES: CASES AND RESOURCES FOR A DIVERSE AMERICA (2nd ed. 2007); RACE LAW STORIES (Rachel F. Moran & Devon Wayne Carbado eds., 2008). For the application of critical race theory outside the legal academy, see Ricky Lee Allen, *Whiteness and Critical Pedagogy*, in CRITICAL PEDAGOGY AND RACE 53 (Zeus Leonardo ed., 2005); Tara J. Yosso & Daniel G. Solórzano, *Conceptualizing a Critical Race Theory in Sociology*, in BLACKWELL COMPANION TO SOCIAL INEQUALITIES 117 (Mary Romers & Eric Margolis eds., 2005).
  4. For review of the literature, see Gómez, *supra* note 1.
  5. See ARIELA J. GROSS, WHAT BLOOD WON'T TELL: A HISTORY OF RACE ON TRIAL IN AMERICA (2008); IAN F. HANEY LÓPEZ, WHITE BY LAW: THE LEGAL CONSTRUCTION OF RACE (1996); RACHEL F. MORAN, INTERRACIAL INTIMACY: THE REGULATION OF RACE AND ROMANCE (2001); Bethany R. Berger, *Red: Racism and the American Indian*, 56 UCLA L. REV. 591 (2009); Kitty Calavita, *The Paradoxes of Race, Class, Identity and 'Passing': Enforcing the Chinese Exclusion Acts, 1882-1910*, 25 LAW & SOC. INQUIRY 1 (2000); Michael A. Elliott, *Telling the Difference: Nineteenth-Century Legal Narratives of Racial Taxonomy*, 24 LAW & SOC. INQUIRY 611 (1999); Katherine M. Franke, *Becoming a Citizen: Reconstruction Era Regulation of African American Marriages*, 11 YALE J.L. & HUMAN. 251 (1999); Laura E. Gómez, *Race, Colonialism and Criminal Law: Mexicans and the American Criminal Justice System in Territorial New Mexico*, 34 LAW & SOC'Y REV. 1129 (2000); Philip Goodman, *"It's Just Black, White, or Hispanic": An Observational Study of Racializing Moves in California's Segregated Prison Reception Centers*, 42 LAW & SOC'Y REV. 735 (2008); Neil Gotanda, *Comparative Racialization: Racial Profiling and the Case of Wen Ho Lee*, 47 UCLA L. REV. 1689 (2000); Neil Gotanda, *A Critique of "Our Constitution is Color-Blind"*, 44 STAN. L. REV. 1 (1991); Cheryl

- I. Harris, *Whiteness as Property*, 106 HARV. L. REV. 1707 (1993); Tanya Katerí Hernández, *Sexual Harassment and Racial Disparity: The Mutual Construction of Race and Gender*, 4 J. GENDER RACE & JUST. 183 (2001); Jerry Kang, *Cyber-Race*, 113 HARV. L. REV. 1130 (2000); Kenneth W. Mack, *Law, Society, Identity, and the Making of the Jim Crow South: Travel and Segregation on Tennessee Railroads, 1875-1905*, 24 LAW & SOC. INQUIRY 377 (1999); George A. Martinez, *The Legal Construction of Race: Mexican-Americans and Whiteness*, 2 HARV. LATINO L. REV. 321 (1997); Michelle McKinley, *Fractional Freedoms: Slavery, Legal Activism, and Ecclesiastical Courts in Colonial Lima, 1593-1689*, 28 LAW & HIST. REV. 749 (2010); Elizabeth Mertz, *The Uses of History: Language, Ideology and Law in the United States and South Africa*, 22 LAW & SOC'Y REV. 661 (1998); Daniel J. Sharfstein, *The Secret History of Race in the United States*, 112 YALE L.J. 1473 (2003); Deenesh Sohoni, *Unsuitable Suitors: Anti-Miscegenation Laws, Naturalization Laws, and the Construction of Asian Identities*, 41 LAW & SOC'Y REV. 587 (2007).
6. See, e.g., LAURA E. GÓMEZ, *MISCONCEIVING MOTHERS: LEGISLATORS, PROSECUTORS, AND THE POLITICS OF PRENATAL DRUG EXPOSURE* (1997); MAE M. NGAI, *IMPOSSIBLE SUBJECTS: ILLEGAL ALIENS AND THE MAKING OF MODERN AMERICA* (2004); SHERENE H. RAZACK, *CASTING OUT: THE EVICTION OF MUSLIMS FROM WESTERN LAW AND POLITICS* (2008); JOHN TEHRANIAN, *WHITEWASHED: AMERICA'S INVISIBLE MIDDLE EASTERN MINORITY* (2009); ERIC K. YAMAMOTO ET AL., *RACE, RIGHTS AND REPARATION: LAW AND THE JAPANESE AMERICAN INTERNMENT* (2001); Devon W. Carbado, *(E)racing the Fourth Amendment*, 100 MICH. L. REV. 946 (2002); Leti Volpp, *Impossible Subjects: Illegal Aliens and Alien Citizens*, 103 MICH. L. REV. 1595 (2005); Leti Volpp, *The Citizen and the Terrorist*, 49 UCLA L. REV. 1574 (2002).
  7. Ian Haney López, *Introduction*, in *RACE, LAW AND SOCIETY*, *supra* note 1, at xi, xviii; see also Gómez, *supra* note 1, at 462.
  8. For example, employment discrimination and immigration law are categories of legal doctrine and policy that deeply implicate race, but I did not include books that primarily trace the evolution of legal doctrine in a particular historical period. See, e.g., ANDREW GYORY, *CLOSING THE GATE: RACE, POLITICS, AND THE CHINESE EXCLUSION ACT* (1998); PAUL D. MORENO, *FROM DIRECT ACTION TO AFFIRMATIVE ACTION: FAIR EMPLOYMENT LAW AND POLICY IN AMERICA, 1933-1972* (1997); LUCY E. SALYER, *LAW HARSH AS TIGERS: CHINESE IMMIGRANTS AND THE SHAPING OF MODERN IMMIGRATION LAW* (1995).
  9. I did not include books that consider law and race from this perspective but that are not rooted in a thick description of a specific place or region. See, e.g., RISA L. GOLUBOFF, *THE LOST PROMISE OF CIVIL RIGHTS* (2007); HANEY LÓPEZ, *supra* note 5; PASCOE, *supra* note 1.
  10. See MICHAEL OMI & HOWARD WINANT, *RACIAL FORMATION IN THE UNITED STATES: FROM THE 1960S TO THE 1990S* (2d ed. 1994); HOWARD WINANT, *THE WORLD IS A GHETTO: RACE AND DEMOCRACY SINCE WORLD WAR II* (2001); Samuel R. Lucas & Marcel Paret, *Law, Race, and Education in the United States*, 1 ANN. REV. L. & SOC. SCI. 203 (2005); Leith Mullings, *Interrogating Racism: Toward an Antiracist Anthropology*, 34 ANN. REV. ANTHROPOL. 667 (2005).
  11. See Devon W. Carbado & Rachel F. Moran, *Introduction*, in *RACE LAW STORIES*, *supra* note 3.

12. Ann Morning, *Toward a Sociology of Racial Conceptualization for the 21<sup>st</sup> Century*, 87 SOC. FORCES 1167, 1171 (2009).
13. EDUARDO BONILLA-SILVA, RACISM WITHOUT RACISTS: COLOR-BLIND RACISM AND THE PERSISTENCE OF RACIAL INEQUALITY IN THE UNITED STATES (2003).
14. See Brackette F. Williams, *A Class Act: Anthropology and the Race to Nation Across Ethnic Terrain*, 18 ANN. REV. ANTHROPOL. 401, 402 (1989).
15. See Gotanda, *A Critique of "Our Constitution is Color-Blind"*, *supra* note 5.
16. See Morning, *supra* note 12.
17. OMI & WINANT, *supra* note 10, at 55.
18. Lucas & Paret, *supra* note 10, at 203–04.
19. See Vilna Bashi, *Racial Categories Matter Because Racial Hierarchies Matter: A Commentary*, 21 ETHNIC & RACIAL STUDIES 959 (1998).
20. See Faye V. Harrison, *The Persistent Power of "Race" in the Cultural and Political Economy of Racism*, 24 ANN. REV. ANTHROPOL. 47, 48 (1995); Morning, *supra* note 12, at 1173.
21. STEPHEN CORNELL & DOUGLAS HARTMANN, ETHNICITY AND RACE: MAKING IDENTITIES IN A CHANGING WORLD 25 (1998).
22. See Morning, *supra* note 12; Am. Anthropological Ass'n, *Statement on 'Race'*, <http://www.aaanet.org/stmts/racepp.htm> (May 17, 1998); Am. Sociological Ass'n, *The Importance of Collecting Data and Doing Social Scientific Research on Race* (2003), available at [http://www2.asanet.org/media/asa\\_race\\_statement.pdf](http://www2.asanet.org/media/asa_race_statement.pdf)<http://www.aaanet.org/stmts/racepp.htm>.
23. Am. Anthropological Ass'n, *supra* note 22; see also Howard Winant, *Race and Race Theory*, 26 ANN. REV. SOC. 169 (2000); Am. Sociological Ass'n, *supra* note 22.
24. Mullings, *supra* note 10, at 674; see also Am. Anthropological Ass'n, *supra* note 22; Am. Sociological Ass'n, *supra* note 22.
25. See CORNELL & HARTMANN, *supra* note 21; OMI & WINANT, *supra* note 10.
26. Am. Sociological Ass'n, *supra* note 22.
27. OMI & WINANT, *supra* note 10.
28. OMI & WINANT, *supra* note 10, at 60; see also Mullings, *supra* note 10.
29. RENISA MAWANI, COLONIAL PROXIMITIES: CROSSRACIAL ENCOUNTERS AND JURIDICAL TRUTHS IN BRITISH COLUMBIA, 1871-1921 (2009).
30. *Id.* at 5.
31. *Id.* at 66.
32. *Id.* at 68.
33. *Id.* at 87–90.
34. *Id.* at 87.
35. *Id.* at 101–02, 108.
36. *Id.* at 109.
37. *Id.* at 109–10, 119.
38. GÓMEZ, *supra* note 6.
39. *Id.* at 81–98.



40. *Id.*; see also Laura E. Gómez, *Opposite One-Drop Rules: Mexican Americans, African Americans, and the Need to Reconceive Turn-of-the-Twentieth-Century Race Relations*, in *HOW THE UNITED STATES RACIALIZES LATINOS: WHITE HEGEMONY AND ITS CONSEQUENCES* (José A. Cobas et al. eds., 2009).
41. IAN F. HANEY LÓPEZ, *RACISM ON TRIAL: THE CHICANO FIGHT FOR JUSTICE* (2003).
42. See Mary Romero, *Brown Is Beautiful*, 39 *LAW & SOC'Y REV.* 211 (2005).
43. HANEY LÓPEZ, *supra* note 41, at 6.
44. Sociologist Mary Romero has criticized the idea of common sense racism as overly psychological and distracting away from the role of the state and political economy in perpetuating racism (issues better addressed by the older concept of institutional racism, she says). Romero, *supra* note 42, at 225–27.
45. HANEY LÓPEZ, *supra* note 41, at 113–27.
46. PEM DAVIDSON BUCK, *WORKED TO THE BONE: RACE, CLASS, POWER, & PRIVILEGE IN KENTUCKY* (2001). For critiques of the book, see Eric Arnesen, Book Review, 66 *INT'L LAB. & WORKING-CLASS HIST.* 218 (2004) (reviewing BUCK, *supra*); Penny Messinger, Whiteness Study Plumbs Kentucky, *H-NET REV.* (Aug. 2002) (reviewing BUCK, *supra*), <http://www.h-net.org/reviews/showrev.php?id=6597>.
47. BUCK, *supra* note 46, at 30–31.
48. *Id.* at 31.
49. *Id.* at 32.
50. *Id.* at 41.
51. *Id.* at 33.
52. JULIE NOVKOV, *RACIAL UNION: LAW, INTIMACY, AND THE WHITE STATE IN ALABAMA, 1865-1954* (2008).
53. *Id.* at 4.
54. *Id.* at 16.
55. *Id.* at 72–74.
56. *Id.* at 142.
57. MOON-HO JUNG, *COOLIES AND CANE: RACE, LABOR, AND SUGAR IN THE AGE OF EMANCIPATION* (2006).
58. *Id.*; Arthur L. Stinchcombe, *Cultures of Discipline: Law Teaches Hawaii to Become a Colony*, 28 *LAW & SOC. INQUIRY* 591, 601 (2003).
59. JUNG, *supra* note 57, at 190.
60. *Id.* at 216–17.
61. See Erika Lee, Book Review, 41 *J. SOC. HIST.* 1081 (2008) (reviewing JUNG, *supra* note 57).
62. VIRGINIA R. DOMÍNGUEZ, *WHITE BY DEFINITION: SOCIAL CLASSIFICATION IN CREOLE LOUISIANA* (1986).
63. See PASCOE, *supra* note 1, at 11.
64. DOMÍNGUEZ, *supra* note 62, at 31–32.
65. *Id.* at 36–37. For a similar pattern involving Indians in 1920s Virginia, see HELEN C. ROUNDTREE, *POCAHONTAS'S PEOPLE: THE POWHATAN INDIANS OF VIRGINIA THROUGH FOUR CENTURIES* (1990).

66. DOMÍNGUEZ, *supra* note 62, at 52.
67. PASCOE, *supra* note 1, at 133.
68. SALLY ENGLE MERRY, COLONIZING HAWAII: THE CULTURAL POWER OF LAW (2000).
69. J. KEHAULANI KAUANUI, HAWAIIAN BLOOD: COLONIALISM AND THE POLITICS OF SOVEREIGNTY AND INDIGENEITY (2008).
70. *Id.* at 91.
71. *Id.* at 81–82.
72. *Id.* at 91.
73. *Id.* at 94–96.
74. *Id.* at 41–42. This may be the least convincing aspect of the book. For a critique that Kauanui’s contemporary political agenda has influenced her historical analysis, see Emma Kowal, *Of Transgression, Purification and Indigenous Scholarship*, 10 ASIA PAC. J. ANTHROPOLOGY 231 (2009).
75. KAUANUI, *supra* note 69, at 8, 70.





**Douglas Lichtman**  
Professor of Law

Douglas Lichtman focuses his teaching and research on topics relating to law and technology. His areas of specialty include patent and copyright law, telecommunications regulation, and information strategy and economics.

A graduate of Yale Law School, Professor Lichtman joined the faculty of UCLA School of Law in 2007 after a tenured teaching career at the University of Chicago. His work has been featured in numerous journals, including the *Journal of Law & Economics*, the *Journal of Legal Studies*, the *Yale Law Journal* and the *Harvard Business Review*. He co-authored *Telecommunications Law and Policy*, a textbook that investigates the federal regulatory regime applicable to broadcast television, cable television, radio, telephony and the Internet. He also regularly writes in the popular press, with recent pieces appearing in the *Los Angeles Times* and the policy magazine, *Regulation*.



# UNDERSTANDING THE RAND COMMITMENT\*

Douglas Lichtman

In a typical agreement between a buyer and a seller, price is one of the central terms specified in the deal. Yet, in a surprisingly large number of technology agreements, patent holders today are choosing to leave out that critical detail. Instead, in these modern agreements, patent holders adopt as their pricing term only a commitment to later price at a “reasonable and nondiscriminatory” rate. This RAND commitment has been used in patent deals covering everything from 3G cellular communication to DVD video playback. But why are firms adopting it? And how should courts interpret its language? In this essay, I take up these questions, considering the purpose behind this type of price ambiguity and ultimately arguing that, at its core, the RAND commitment is most likely a pro-competitive mechanism primarily designed to guide courts away from patent law’s conventional—and here largely inappropriate—damages regime.

ABSTRACT

**P**rior to adopting a technical standard, standard-setting organizations typically endeavor either to make sure that the standard does not infringe any patent rights or to clear the necessary permissions. The task is a difficult one. The protocol that governs how information is stored on DVD-R media, for example, is known to implicate at least 342 different patents. The encoding, decoding, and transmission protocols relevant to just one type of cellular telephony touch well over 1,000.

I. INTRODUCTION

Those large numbers are problematic because it takes substantial time and money to evaluate a patent. To do the job right, consensus would have to be achieved as to whether the patent is valid, whether it covers a truly essential aspect of the standard at issue, and exactly how much the patent contributes as compared to next-best alternatives. Worse, all of that would need to be done in a context where patent holders have strong incentives to exaggerate, where information about unpublished patent applications is understandably hard to come by, and where there will often rage an independently contentious debate over nonpatent issues like the specifics of what should and should not be included in the standard.

Thus enters the RAND commitment.<sup>2</sup> Instead of undertaking the difficult task of evaluating asserted patents, most standard-setting organizations simply keep a running list of patents that have been asserted to be relevant by one or another patent-holding participant. Those participants are then required to agree that ultimately, they will make available to the public, on “reasonable and nondiscriminatory terms,” any truly essential patent. The need for careful patent analysis is thereby diminished. If a given patent turns out to be irrelevant, no one will need a license for its use anyway. But if a given patent turns out to be essential, at least the relevant patent holder has promised to license at a reasonable and nondiscriminatory rate.

Hidden in that simple solution, of course, is enormous complexity. Is the RAND commitment a license, such that firms can go ahead and implement the technology subject only to a later obligation to negotiate the price? Is it a promise to license, which would mean that implementing firms in fact have no right to use the patented technology until they cut a specific deal? And what happens if, as seems enormously likely, an essential patent holder ultimately thinks one price is reasonable, whereas the implementing firms think a much lower number is appropriate? Is that dispute a contract dispute, litigated using traditional contract damages measures, or a patent dispute, meaning that the patent system’s damages regime controls?

In this excerpt, I set out to answer some of these fundamental questions. I begin in Part II by articulating four reasons why firms involved in the standard-setting process use RAND rather than explicitly negotiating price. My main point is that participating firms are attempting to delay price negotiation, but in a way that does not distort that negotiation when it ultimately does take place. In Part III, I build on that insight to advance one primary contention. I argue that triple damages and injunctive relief are both inconsistent with the purpose of RAND, and thus patent holders should not be allowed to invoke those traditional patent remedies in the context of a RAND dispute. In Part IV, I consider alternative interpretations of RAND that would leave traditional patent law remedies more fully intact. Part V then briefly concludes.

## II. WHY RAND?

I want to start by thinking about RAND from a purely business perspective. That is, for the moment I want to put aside patent, contract, and antitrust law, and I want to focus instead on the business question of why firms might prefer the ambiguous RAND commitment over a more conventional, explicit pricing term. The answer is not immediately obvious. After all, in most business settings, buyers very much want to know the prices associated with competing options so that they can ultimately make tradeoffs between price and quality. Yet, in standard setting, that norm has been largely abandoned. Why?

One reason is that intricate negotiations over patent validity and patent value would take an enormous amount of time. To work through a process where dozens of companies would debate the merits and worth of hundreds of patents would take years. Worse, were consensus not achieved, litigation would run yet more time off the clock, with substantial time lost first at the district court and then on appeal. One charm of the RAND commitment for participants and the public alike, then, is that RAND allows technological implementation to move forward while the parties in parallel work out legal and financial details.

A second and related reason that firms opt for the RAND commitment is that standard setting is a process run by engineers, not lawyers. A technology firm like Microsoft or Dolby can easily be involved in dozens of standard-setting processes at the same time. To send to each of those bodies not only the obviously necessary engineers but also an army of lawyers, business executives, and pricing specialists would be enormously expensive. The RAND commitment thus simplifies the conversation, allowing the engineers alone to run the show until the technical details are fully selected and documented.

A third reason that firms choose RAND is that many new technologies flop. Digital Audio Tape (DAT) technology for a time looked like it was going to be so important that Congress passed a set of laws specifically regulating its sale and use.<sup>3</sup> Oops. Similarly, from 2002 through 2008, the HD-DVD standard was backed by industry heavyweights including Toshiba, Sanyo and NEC. Today, HD-DVD technology, too, is just a footnote in history. Thank goodness, then, that EMI, Sony BMG, Universal Music Group, Warner Music, Sony, Toshiba, Sanyo, NEC, Paramount, Hewlett Packard, Microsoft, Apple and their peers did not each invest a fortune vetting those patent situations. They would have negotiated detailed terms for a group of patents that turned out to have little actual commercial value.

A fourth reason that firms opt for RAND—and this is in essence a more general version of reason three—is that RAND allows implementing firms to wait for additional information before they commit to a specific royalty structure. When Al Gore invented the Internet,<sup>4</sup> no one really understood the impact those protocols would have on commerce, culture, and communication. Much the same, when the now-familiar 2G wireless standard was first promulgated, even that technology's strongest proponents could not have foreseen the degree to which cell phone usage would permeate both work and play. Financial arrangements will often be more efficient in the long run if their details can be negotiated after the negotiating parties more fully understand how the technology at issue is going to be used and by whom. The RAND commitment delays pricing negotiations and thereby allows at least some of that information to be included in the ultimate royalty negotiation.



None of this is meant to imply that a RAND-like approach is inevitable, or even that RAND is clearly the right way to go. My points about delay, failure, and the desirable absence of lawyers are all general arguments that could apply with comparable force in other settings. Businesses are routinely forced to delay the launch of products and services while they lawyer up their relationships and negotiate elaborate deals. And yes, many of those complicated deals ultimately prove worthless because consumers reject the resultant offering. Just the same, in many settings, delay would allow beneficial information to come to light—for instance, information about consumer preferences and the pace of market adoption. But businesses all the time enter into contracts anyway, allocating risk without knowing exactly what will happen next and sometimes making their various obligations contingent on future events. So, while these are all real problems to be sure, RAND is neither the only nor a necessary way to overcome them.

Moreover, RAND has some real drawbacks. Consider, for instance, the fact that the RAND commitment separates the negotiation over the details of a technology from the negotiation over its cost. At my house, that would be an obvious and unacceptable blunder. I can only imagine what my wife would say were I to choose an expensive piece of home electronic equipment—say, a new flat-screen television—without having an estimate of what that hardware and its accessories would ultimately cost. Yet sophisticated companies like Nokia, Ericsson, Nortel, Sony, InterDigital, Texas Instruments, Alcatel, DirecTV, and NEC did exactly that when they hammered out the details of the recently launched Long Term Evolution (LTE) telecommunications standard. They made hundreds of nuanced choices about how LTE-compliant devices would decode, encode, and transmit data, but they did so without any real understanding of what the various options would cost in terms of patent fees. That sort of economic blindness is par for the course in the standard-setting process. But it is still jarring, and it still represents a real downside to the RAND approach.

As I will discuss more fully later in the essay, RAND has another significant drawback as well: It forces courts to take a more active role when it comes to pricing patents. Judges and juries are unlikely to be very good at valuing patented inventions. This would be true in even a simple case where a single patent was at play, but it is all the more true in the context of standard setting, where the value of any one patent has to be judged in light of hundreds or even thousands of other necessary patent rights. An explicit pricing regime would address this difficulty: Standard-setting participants would negotiate each patent's appropriate price, and courts would be asked only to enforce the agreed-upon deals. The RAND commitment, by contrast, puts courts center stage. If the parties cannot agree as to what "reasonable" means, a judge or a jury will ultimately have to wade through the evidence and pick a

number. Moreover, even if the parties in the end agree on what “reasonable” means, their agreement will unavoidably have been influenced by what each party expected a court would do if agreement had not been reached. RAND, then, takes a task that courts perform poorly and makes that task a central driver in the ultimate economic interaction.

All that said, however, the RAND commitment is, as of this writing, widely used in settings where a large number of patents are plausibly implicated by a given technological standard. It might be a good idea. It might be a terrible one. But RAND is today pervasive, and as such it is inevitable that courts will at some point have to interpret its meaning. It is to that topic that I now turn.

**I**t is something of an outrage that the language of the RAND commitment offers so little guidance as to its proper interpretation. What is “reasonable,” and how does use of that word compare to patent law’s use in the context of “reasonable” royalties? The standard RAND clause does not say. Does “nondiscriminatory” mean that prices must be the same across the board, or does it mean that some degree of price differentiation is fine, but differences keyed to certain distasteful characteristics—*discrimination*—are verboten? I suspect the latter, but again, the standard clause does not elaborate. Bad enough if this sort of ambiguity had shown up in some private contract governing an interaction of modest economic and cultural import. But the RAND commitment governs patent rights in a breathtaking array of familiar industries and technologies, and in the end it will serve to allocate multiple billions of dollars between and among major firms.

### III. INTERPRETING RAND

Courts might be tempted to punish this ambiguity by interpreting the commitment to be meaningless, nonbinding, or otherwise ineffective. Such an approach would pressure firms in the future to draft contract language with more precision, in essence implementing the Ayres/Gertner “penalty default” concept.<sup>5</sup> That, however, would be a mistake. Yes, looking forward, standard-setting organizations ought to supplement the traditional RAND language with more specific information about exactly what the clause means. And one can easily imagine a future RAND clause that reads, say, “reasonable and nondiscriminatory, by which we mean that the patent holder has no right to an injunction” or “reasonable and nondiscriminatory, by which we mean to endorse patent law’s traditional *Georgia Pacific* factors.”<sup>6</sup> But for the RAND commitments already in place, courts are stuck with only two choices: interpret the clause in light of its likely purpose, or strip it of meaning and in that way throw into turmoil the economics that undergird countless important consumer technologies. I myself favor the former approach, and so here I consider how RAND likely ought to be read from both an economic and a patent law perspective.

A. The Economic  
Perspective

From an economic perspective, the purpose of the RAND commitment is to ensure that patent holders are not able to earn exaggerated royalties merely because price negotiations have been delayed. Without some sort of pricing commitment, this is exactly what would happen. The price for any technology included in the standard would go up simply because it was chosen. That is emphatically not the point of RAND. Standard-setting participants defer pricing negotiations because they want more information, or because they want to implement the relevant standard more quickly, or because they want to minimize upfront costs. But it seems implausible to think that standard-setting participants opt for RAND in order to randomly and artificially increase each patent holder's ultimate leverage.

To see this point more fully, consider a situation in which two comparable technologies are vying for inclusion in a given standard: Dolby's high-fidelity audio compression codec on the one hand and DTS's rival audio compression technology on the other. Were prices being negotiated at the time of the selection, participants in the standard-setting process would compare the Dolby and DTS approaches. They would identify advantages and disadvantages, and they would ultimately offer the winner a price that reflected its marginal value as compared to the unsuccessful alternative. If the winning patent holder were to hold out for more, standard-setting participants would presumably threaten to switch to the second-best technology. Ultimately, a competitive bidding process would typically yield something close to the efficient price.<sup>7</sup>

Now consider what would happen if, instead of negotiating at the time of selection, standard-setting participants were to wait and negotiate a few years later. Two important considerations would by then have changed. First, a given licensee would by that point likely have made some technology-specific investments. The firm would have designed its products. It would have built manufacturing facilities. It would have made commitments to buy components from its suppliers. And it would have promised relevant functionality to downstream customers. A patent holder would be able to take advantage of all of those commitments, demanding a royalty that reflected not only the value of the patented technology as compared to next-best alternatives, but also the value that this licensee would place on avoiding disruptions to its already-made investments.

Second, even if a particular would-be licensee had not made patent-specific investments, its peers would have—and that triggers a similar dynamic. Consider standards with respect to driving. Before driving norms were established, the value of "driving on the left" was roughly equal to the value of "driving on the right." Everyone surely agreed that all the drivers in any particular region ought to adopt the same default rule, but the choice between the two was likely a draw, and thus

patents related to either one would have been of similar value. Once a great deal of traffic had opted for the right, however, the economics changed. A patent related to the idea of driving on the left was worth very little. A patent related to the idea of driving on the right was worth a fortune. The change had nothing to do with the relative merits of these two technologies. It was just an example of a more general phenomenon associated with standardization: Patents related to a chosen standard increase in value as more people adopt the standard. This is then another problem with delayed patent negotiations. Patent holders who negotiate after standardization are able to charge prices that reflect the now-realized network value, in essence charging licensees for the fact that other licensees are already committed.

From an economic perspective, then, the important work of the RAND commitment is to minimize these economic distortions. Participants in the standard-setting process might well intend to pay patent holders the royalties they would have gotten had sufficient time, cash, and predictive information been available so as to enable complete negotiation *ex ante*. And participants in the standard-setting process might also intend to pay each patent holder a bit extra as thanks for that patent holder's willingness to delay the price negotiation and in that way reduce the costs associated with the standard-setting process. But there is no reason to believe that standard-setting participants also mean to allow patent holders to hold hostage each participant's standard-specific investments, or to allow patent holders to arrogate to themselves the value created from the group's standardization effort.

From a patent perspective, the purpose of the RAND commitment is to reject patent law's default damages regime. Patent law, it turns out, does not award "reasonable" royalties. Quite the opposite: When a court decides that a valid patent has been infringed, the court typically imposes a remedy the net value of which clearly exceeds the value of any deal the parties would have made had they negotiated a license prior to the infringement. This exaggeration is explicit and intentional. But, with just one exception, the reasons for it do not apply in the RAND context.

## B. The Patent Law Perspective

One reason patent law intentionally exaggerates is that exaggeration encourages private parties to negotiate rather than litigate. The mechanism here is obvious: An infringer who knows that litigation will yield exaggerated liability has strong incentives to avoid litigation. This is good public policy primarily because judges and juries are not well equipped to value patented inventions. True, they might be able to make educated guesses based on the evidence presented, and they might even be as likely to overestimate as they are to underestimate. But in any specific dispute, there is little reason to believe that judges and juries will come up with even a remotely accurate estimate for the value of the patent at issue.

### 1. The General Case for Exaggeration

Private negotiation, by contrast, can be reliable so long as private negotiation takes place before the would-be infringer has made any investments that are specifically tied to the patented technology.<sup>8</sup> In those circumstances, the negotiation between the patent holder and a potential infringer resembles a competitive interaction. The patent holder can ask for a high starting price; the potential infringer can counter by pointing to potential substitute technologies; and ultimately the process should yield a price that accurately reflects the marginal advantages of the patented technology.

Exaggeration thus can serve a useful purpose. Whenever infringers are able to negotiate prior to making any patent-specific investments, exaggeration helpfully increases their desire to do so. The net result is a patent system where patents are more likely to be priced in the private market, and courts can therefore avoid the difficult task of valuing patented inventions themselves.

The second justification for patent law exaggeration derives from the concern that without exaggeration, infringers would have a strong incentive to hide their illegal activity instead of addressing it. My remarks thus far already speak to this concern in part. My first justification for exaggeration was that it creates an incentive for an infringer to negotiate prior to making patent-specific investments. Obviously, to negotiate, an infringer will have to identify himself to the patent holder, and thus an incentive to negotiate is simultaneously a disincentive to hide. But hiding is a bigger issue than just that. Consider, for instance, an infringer who neither knew nor could have known about a patent prior to making patent-specific investments. If this infringer later discovers the patent, he will be reluctant at that point to contact the patent holder and negotiate because then the patent holder will hold hostage the infringer's already-made investments. Worse, this infringer will have an affirmative incentive to keep quiet. After all, the patent holder might never even notice the infringement, and hence, if the infringer keeps his head down, he might never have to pay. The patent system on these facts faces a real challenge: The system needs to create an incentive for the infringer to identify himself, but at the same time it needs to protect him from an undesirable hostage situation.<sup>9</sup>

Exaggerated damages can solve this problem. Where there is evidence that the licensee reasonably could have stepped forward but chose not to, patent law can punish that choice by exaggerating. The infringer would suffer because of his decision to hide, and in the long run that would encourage infringers in similar situations to step forward. By contrast, in cases where the licensee does step forward, patent law can promise to take its thumb off the scale, calculating royalties with an eye toward the royalty the parties would have struck had they been able to strike a deal before the infringer first invested. The parties would then hopefully foresee that evenhanded result and negotiate in its shadow. But the key point is that the backstop to their negotiation would be the threat of a truly reasonable

royalty and not the threat of either a hostage-taking situation or exaggerated court remedies.<sup>10</sup>

The third justification for patent law exaggeration is simply this: Exaggeration is the way the patent system accounts for changes in patent uncertainty. Prior to litigation, there is almost always some uncertainty as to whether the patent at issue is valid, whether the patent at issue actually has been infringed, or both. That is, the accused infringer might plausibly argue that the patent should never have been issued, and the accused infringer might similarly argue that its technology is not covered by the patent's claims. Litigation resolves the uncertainty. Thus, when a patent holder prevails, the damages awarded naturally are higher than the royalties the parties would have negotiated prior to verdict. Negotiated royalties reflect uncertainty; court-determined royalties do not. Think of it this way: If prior to litigation a patent holder and a would-be licensee both agree that there is a 50 percent chance that the asserted patent is invalid, their private deal would reflect those doubts. The licensee would demand a discount as compared to a sure-thing royalty, and the patent holder would accept that discount in order to avoid the risk of a bad outcome. If that patent holder ends up successfully litigating the issue, however, the resulting court-ordered royalty should no longer reflect that 50 percent discount. Had the patent holder lost the case, he would have earned nothing. Given that he won, he should correspondingly earn the undiscounted award. Intuitively, that's what it means to take the risk of actually litigating the issues.

There are many patent doctrines that implement patent law's exaggeration approach. These doctrines largely track the policy intuitions sketched above. An obvious example is the rule with respect to willful infringement. Under current law, if an infringer knew or should have known about a patent, and if that patent's validity and relevance is objectively clear, then the court can use its discretion to award up to triple damages.<sup>11</sup> The logic is that, on those facts, the infringer was presumably in a great position to seek out the patent holder and negotiate a license prior to the infringement. The infringer knew or should have known about the patent, and the infringer should have paid up because the patent was so clearly valid and infringed. Courts therefore are empowered to triple the damages, thereby encouraging negotiation in similar future situations.<sup>12</sup>

Another way patent law implements its exaggerated damages regime is through the system's willingness to issue injunctions against future infringement. When a patent holder wins an infringement case, the patent holder typically requests not only cash for infringement that has already occurred, but also an injunction barring future infringement. The idea is that the injunction will force the infringer to negotiate with the patent holder, and thus the private parties will set their own forward-looking

## 2. How Patent Law Exaggerates

royalty rather than relying on the judge or the jury to do so.<sup>13</sup> That negotiated royalty will be exaggerated, however, because of the sunk-cost problem discussed earlier. A firm that has already begun to use what turns out to be an infringing technology will typically also already have made investments specific to that technology. The firm's manufacturing facilities will already be tailored to produce the infringing component. The firm's contracts with its suppliers and its customers will already be tethered to that and not some other technical approach. As a result, the infringer will be willing to overpay for the technology, paying the intrinsic value of the technology as compared to its next-best substitute plus a kicker that reflects the savings associated with not having to change its production facilities or in other ways disrupt existing business relationships and practices.<sup>14</sup>

A third way patent law exaggerates is through a false assumption that is nevertheless routinely employed in damages analysis. When a patent holder sues for damages, he can request that a "reasonable royalty" be determined through what is known as a "hypothetical negotiation" framework. As explained by the courts, the hypothetical negotiation simulates the conversation that the infringer and the patent holder would have had if they had negotiated prior to the first act of infringement.<sup>15</sup> The courts endeavor to run the simulation accurately, even going so far as to consider only information that was actually available at the time the negotiation would have occurred. But courts then make one initially surprising move: They assume that both parties involved in the negotiation believe the patent to be valid and infringed.<sup>16</sup> In reality, of course, there would almost always have been doubt. But this false assumption is made for the reason I discussed above: A patent holder whose patent survives litigation must be compensated for having incurred that risk, and so the royalty calculation made after verdict must build patent validity and infringement into the math.

### 3. Exaggeration and RAND

Reasonable minds can disagree over whether these various exaggeration techniques are in the end an effective way to address the public policy concerns that justify them. My own view is that these doctrines need to be more explicitly tied to their underlying policy goals because today patent holders seem to abuse these rules by invoking them in situations where they ought not apply. For the purposes of this essay, however, my point is more narrow. Whatever its merits in general, exaggeration like this is for the most part inappropriate as applied to patents covered by the RAND commitment.

The first policy consideration, the idea of encouraging negotiation prior to investment, is clearly inapt. The whole purpose of the RAND commitment is to allow patent holders and would-be infringers to delay negotiation. Yes, one consequence of that delay is that infringers and patent holders both miss out on

the chance to negotiate prior to investment. And that is a real cost to the system: Private negotiation prior to investment is surely a more accurate means by which to establish patent value than are alternatives like private negotiation after investment or explicit court determination. However, that is the choice RAND embodies. Thus, to the extent the RAND commitment is going to be enforced—and I have argued that it should be, at least with respect to deals already in place—it makes no sense to impose exaggerated damages to punish the very delay RAND set out to accomplish.

The second policy consideration—the concern about undetected infringement—also resonates poorly in the RAND context. Patent holders in conventional settings find it difficult to identify infringers because infringers tend not to speak up. Moreover, infringing products and processes can be hard to reverse-engineer, and that too makes detection by the patent holder difficult at best. In the RAND context, however, these problems are either fully eliminated or substantially reduced. For one thing, participants in the standard-setting process identify themselves publicly. No hiding there. Similarly, firms that produce products or implement processes consistent with a standard also typically self-identify. They label their products as standard-compliant, or the fact of standard compliance is obvious upon even casual inspection. Admittedly, that still leaves some special cases where the infringer’s economic footprint might be so small that it goes unnoticed, or where the use of a particular standard is not evident on the face of the relevant product or process. Still, for the most part, detection is not a significant problem in the typical RAND setting, and hence, when interpreting RAND, damages exaggeration cannot be readily justified on that ground.

That leaves only the third policy consideration, the one keyed to patent uncertainty—and good thing. Patent uncertainty in the context of the RAND commitment plays out exactly the same way as it plays out in conventional patent settings. Prior to verdict, a potential licensee will offer a royalty that discounts for the fact that the patent rights are uncertain. After verdict, however, that uncertainty is resolved, and the royalty should adjust accordingly. Indeed, the system would not work otherwise. Suppose that a patent holder and a would-be licensee both thought that there was a 25 percent chance that the patent at issue was valid. If at the end of litigation the court awarded only 25 percent of the sure-thing royalty, the entire negotiation dynamic would unravel. Prior to verdict, the would-be licensee would argue that the patent holder had only a 25 percent chance of winning and, at that, would win only 25 percent of the sure-thing royalty. Thus the licensee would rationally offer a paltry 6.25 percent of patent value, in essence wrongly double-counting the 25 percent discount. Exaggeration is thus still necessary with respect to patent uncertainty, even under RAND.



Where does all that leave us? Injunctive relief primarily serves the first policy consideration, and willful damages primarily serve the second. Both of these measures of patent damages should therefore be off the table in the RAND context. The exaggeration inherent in the reasonable royalty framework, by contrast, primarily serves the third policy consideration, and hence that exaggeration ought to survive.

There are several plausible ways to achieve these outcomes. Courts could interpret RAND as a public commitment that creates a defense of equitable estoppel. Under that estoppel, the patent holder would be deemed to have permanently waived his right to seek triple damages or to ask for injunctive relief, but would otherwise be allowed to invoke patent law's damages regime. Courts could just as well interpret RAND as creating an implied license, with the license rendering moot any claim to injunctive relief or triple damages, but leaving the court with the power to determine the royalty due. I do not mean to choose between these and other options here. I only want to emphasize that no matter what doctrinal lens courts apply, the goal should be to trim patent law's damages regime such that the damages awarded ultimately approximate the royalty the parties would have negotiated prior to standardization plus a kicker for the now-resolved uncertainty.

#### IV. OTHER VIEWS

If my analysis thus far is correct, the RAND commitment is, at its heart, a mechanism by which private parties can delay pricing negotiations without inadvertently skewing the outcome of those negotiations. It is an implicit rejection of the standard patent damages regime (which would very much skew the outcome), and it is less problematic on antitrust grounds than the only obvious alternative: a group-wide, *ex ante*, explicit conversation about price.

Where, then, might other voices disagree? One possible competing interpretation would read the RAND commitment as simply a promise to make a reasonable offer. In this view, a patent holder subject to the RAND commitment must offer a reasonable royalty to each interested licensee; however, if a licensee rejects that offer, all bets are off, and the patent holder is at that point free to exercise its patent rights as if there had been no RAND commitment. I reject this interpretation because its predictable implication is to overcompensate patent holders. After all, there will always be enormous uncertainty over precisely what is and is not a reasonable royalty. Yet if the RAND commitment were read this way, a licensee who in good faith disagreed with a patent holder could pursue that disagreement only by taking the risk of paying double, triple, or more, depending on the details of patent law's exaggeration regime. As a result, licensees would rationally accept

high royalties because the certainty of overpaying slightly would be more attractive than the distinct possibility of overpaying by a multiple.

A variant on this theme would be an interpretation that protects licensees, but only so long as they engage in good faith negotiation. That approach has some charm, but it puts enormous pressure on courts to determine whether a licensee is acting in good faith. I would not object to that if the test were a conservative one, such as a rule that imposed exaggerated damages only in instances where some smoking-gun email made bad faith plain. But I worry about less reliable tests in that they would create the same uncertainty that I sketched above: A licensee acting in good faith would worry that a court might later misconstrue those intentions, and because of that the licensee would knowingly accept a moderately high royalty in order to avoid the risk of an astronomical one.

Another interpretation that would leave the door open to triple damages and injunctive relief is an interpretation under which the patent holder would be required to continually extend a reasonable offer, even after a licensee had previously turned down that offer. The idea here is that the would-be licensee's risk would be capped: The licensee would be exposed to exaggerated damages for as long as the dispute raged, but the licensee could end that exposure at any time by accepting the patent holder's always-open offer. The problem this time is that the period of exaggerated damages could still be significant because patent litigation and its reasonable appeals can easily last years. This would again put an enormous thumb on the scale, pressuring licensees to accept a royalty that is higher than reasonable but not so high so as to warrant the risks of litigation.

In resisting these alternative views, I do not mean to ignore the opposite problem: Under my interpretation, RAND does little to encourage standard-setting participants to negotiate rather than litigate. Remember, patent law solves this problem by threatening to impose exaggerated damages on any infringer who could have negotiated, and should have negotiated, but in fact did not. RAND dismantles that exaggerated remedy regime (for good reason) but then offers nothing to replace it. The upshot might be that patent holders who agree to the RAND commitment will in the end be undercompensated. They will have to either discount their rates in order to lure licensees to the table or incur the costs, risks, and delay of litigation in order to ultimately be paid their due. I favor this distortion over the opposing one, however, because it should be much smaller. The cost of litigation can be large, but it will almost always be rounding error when compared in size to the exaggeration built into patent law's damages regime. Besides, courts in fact offset much of this distortion by requiring adjudged infringers to pay expenses, attorney's fees, interest, and the like.

## V. CONCLUSION

**M**y ambition in this essay was to articulate the reasons why firms in the standard-setting context opt for the RAND commitment rather than explicitly negotiating price, and relatedly to explain what the RAND commitment as a result likely means. My answers are hopefully by this point evident. Firms choose RAND because they want to delay pricing negotiations without inadvertently skewing the outcome of those later deals. As a result, RAND must be interpreted to reject much of the conventional exaggerated patent damages regime. The result is not a perfect framework for patent licensing. However, the possibility of voluntary delay does open the door to significant efficiencies, and surprising as it might be, firms involved in standard-setting obviously believe that those efficiencies outweigh the associated costs.

Am I therefore a fan of the current RAND approach? Hardly. The success of the RAND commitment in the end turns on the ability of a court to calculate a nonexaggerated reasonable royalty. If courts tend to pick royalties that are too high, then private parties negotiating in the shadow of litigation will also choose too-high rates. If courts tend to pick royalties that are too low, then private parties will similarly strike inefficiently modest deals. The only plausible happy story is a story where courts have no predictable or systematic bias. Of that I am enormously skeptical.

More broadly, though, I dislike RAND as it exists today because it could easily be so much more. This is a clause invoked by a veritable who's who of technology and electronics companies. One would think that when firms of that caliber gather together to establish an elaborate agreement about the future of some promising new technology, they would at the same time opt out of the default legal regime and establish their own expert arbitration process to resolve future disputes accurately and at an appropriate pace. But no. That frankly boggles the mind. Where conflicts are a surprise—for instance, any classic tort—the default legal regime is the only option. Prior to a car crash, I cannot by contract agree with the other driver that we will use a more efficient mechanism for allocating fault. But here, sophisticated private parties can foresee conflict, they know that the subject matter of that conflict will be enormously difficult for a lay judge or lay jury to evaluate, and they are already in contact with one another through the standard-setting process. Private dispute resolution should naturally follow.

So, yes, because billions of dollars are today at stake across a host of important industries, courts should interpret RAND with an eye toward the purposes and policies articulated here. And yes, as new RAND commitments are written, standard-setting participants ought to make explicit their repudiation of patent law's exaggerated damages regime. But just as important, standard-setting organizations in the future need to invest in alternative dispute resolution. RAND as it exists today can only be as good as the courts that will enforce it. And with so much inventive activity on the line, that seems hardly good enough.





- \* This is a shortened version of an article originally published as Doug Lichtman, *Understanding the RAND Commitment*, 47 HOUS. L. REV. 1023 (2010). It is excerpted here by permission.
1. Although RAND is the common acronym used in the United States, in Europe the relevant acronym is FRAND, which stands for “fair, reasonable and nondiscriminatory.” For purposes of this Essay, I use the RAND formulation, but everything I say here is equally applicable to the FRAND variation.
  2. I refer here to the now largely irrelevant Audio Home Recording Act of 1992, 17 U.S.C. §§ 1001–1010 (2006).
  3. Just teasing. See Richard S. Whitt, *Adaptive Policymaking: Evolving and Applying Emergent Solutions for U.S. Communications Policy*, 61 FED. COMM. L.J. 483, 548 (2009) (discussing how Vice President Al Gore was wrongly accused of having himself claimed to “invent” the Internet).
  4. Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 YALE L.J. 87, 97 (1989).
  5. See *Georgia-Pacific Corp. v. U.S. Plywood Corp.*, 318 F. Supp. 1116, 1120 (S.D.N.Y. 1970) (articulating the now-famous list of considerations relevant to the calculation of a reasonable patent royalty).
  6. None of this is to rule out the possibility that some technologies are unique and thus would not be substantially constrained by *ex ante* competition. If there were only one way to encrypt data for cellular transmission, for instance, the patent covering that technology would be worth a fortune no matter when its value was determined. Thus, when I refer to a competitive price, I mean only to refer to the price that would be assigned at a time prior to standardization. That price might be low. It might be high. But it would reflect the value inherent in the technology rather than the value created by the decision to standardize.
  7. My words are carefully chosen in the text because stronger statements—say, “private negotiation is reliable so long as the negotiation takes place before the would-be infringer has made any patent-specific investments”—are simply not true. Recall, for instance, my example about the norms related to driving on the right side of the street. In that context, even if an accused infringer had not invested prior to negotiation, the negotiation would still be skewed if a substantial number of other firms had already invested. This is increasingly a problem for the patent system. The patent system’s core assumption about the efficiency of private market transactions might have been valid in a world where products were typically covered by one and only one relevant patent, or a world where network externalities were not of significant economic import. But the modern world does not remotely meet either of those conditions, and that is one of many reasons why the patent system today is troublingly out of kilter.
  8. Although this essay is not the place to explore the issue fully, I cannot help but point out another possible solution to this problem: Deem the infringer’s use of the invention perfectly legal. On the facts I sketch above, the accused infringer independently

invented the patented invention. He did not copy it from the patent holder, he did not know about the patent itself, and he could not have discovered the patent had he engaged in reasonable efforts to find it. In that situation, patent law could in theory declare independent invention to be a complete defense to patent infringement. The patent system exists to reward patent holders for bringing inventions into meaningful public use, either through their own efforts or through proactive licensing. Here, the patent holder did no such thing, and the patent system could take that into account by refusing to recognize infringement.

9. Exaggeration and hostage-taking are different, in that one is calibrated by legal doctrines while the other is random. That is, when patent law is the source of the exaggeration, courts are in theory actively choosing the degree of exaggeration by (say) calibrating the relevant injunction or specifying whether damages ought to be doubled or tripled. In hostage-taking, by contrast, the extent of any overpayment turns on a number of arbitrary factors, including the amount the infringer has already invested in the patented technology and the number of other patent holders who are able to hold hostage that same sunk investment.
10. See *In re Seagate Tech., LLC*, 497 F.3d 1360 (Fed. Cir. 2007) (articulating the modern willfulness doctrine).
11. Discretion is an important part of the doctrine because willful damages are inappropriate in some instances where the formal test is nonetheless met. For instance, sometimes an infringer will hear of an obviously valid patent only after the infringer has made patent-specific investments. If the infringer were at that point to reach out to the patent holder, the patent holder would attempt to hold those investments hostage. Willful damages ought not punish the infringer for turning down that distorted deal. Quite the opposite, when faced with that situation, it is reasonable for an infringer to ask the court for an impartial valuation, rather than simply paying a clearly exaggerated royalty.
12. Sometimes, of course, the patent holder is not interested in negotiating at all, preferring instead to bar infringement and thus limit the number of firms using the invention.
13. The Supreme Court's recent decision in *eBay Inc. v. MercExchange, L.L.C.* helps to check this form of exaggeration because a defendant in this position could convince the court that, under the *eBay* test, an injunction ought not issue. The court would then itself impose a forward-looking royalty that would not take into account patent-specific sunk costs. The law here is not sufficiently developed to know for sure whether that sort of argument will carry the day, but *eBay* at least opens the door. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 391 (2006).
14. See, e.g., *Radio Steel & Mfg. Co. v. MTD Prods., Inc.*, 788 F.2d 1554, 1557 (Fed. Cir. 1986) (explaining the hypothetical negotiation framework). Note that this approach inherently exaggerates, in that the value of the technology would be more accurately represented by the royalty the parties would have chosen had they been able to

negotiate before the infringer made any patent-specific investments. Courts, however, mistakenly run the analysis at the time of the first infringement.

15. See Robert Goldscheider, *The Employment of Licensing Expertise in the Arena of Intellectual Property Litigation*, 36 IDEA 159, 174 (1996).







**Noah D. Zatz**  
Professor of Law

Noah Zatz focuses on employment and labor law, welfare law and antipoverty policy, work/family issues, feminist legal and social theory, and liberal political theory. His teaching and writing address how work structures both inequality and social citizenship in the modern welfare state.

He is particularly committed to training public interest lawyers and to engaging students with law's possibilities both as an instrument of injustice and as a contributor to emancipatory social change. He is an active participant in UCLA's Epstein Public Interest Law & Policy and Critical Race Studies programs.

Before entering legal education, Professor Zatz was awarded a Skadden Fellowship to work at the National Employment Law Project (NELP) in New York City. After graduation from Yale Law School, he clerked for Judge Kimba M. Wood of the U.S. District Court for the Southern District of New York and for Judge Guido Calabresi of the U.S. Court of Appeals for the Second Circuit. He also has been a fellow at Princeton University's Program in Law and Public Affairs, a visiting fellow at the University of New Mexico School of Law and a visiting professor at the University of Chicago Law School.



# SUPPORTING WORKERS BY ACCOUNTING FOR CARE\*

Noah D. Zatz\*\*

In this article, I argue that two pathologies in contemporary antipoverty policy— inadequate childcare assistance and failure to value parental caregiving as work—are flip sides of one coin. That coin is “childcare invisibility.” Childcare invisibility is a ubiquitous feature of “means testing,” a basic technique that targets assistance to low-income households in programs like Temporary Assistance for Needy Families (TANF), Supplemental Nutrition Assistance Program (SNAP, formerly known as Food Stamps), and the Earned Income Tax Credit (EITC). Means testing identifies who is poor, in the sense of lacking the “means” to meet basic needs, and who therefore requires assistance. Unlike food, clothing, shelter, and so on, childcare makes no appearance in the basket of goods and services that these policies use to determine what a household needs to maintain a minimally decent standard of living.<sup>1</sup> Childcare again goes unseen when the government determines whether those needs are being met or, instead, whether the household requires assistance to make ends meet. This absence from the two sides of means-testing— needs and resources—is childcare invisibility.

## INTRODUCTION

Childcare’s invisibility within means testing also implicates childcare’s status as work. This connection arises from two legacies of 1990s welfare reform that made anti-poverty policy into “work” policy as well. First, work requirements became central to benefits for low-income families with children, most notably when TANF replaced the New Deal-era Aid to Families with Dependent Children (AFDC). AFDC initially had not included work requirements, and eventually it became the primary target of efforts to “end welfare as we know it.” Now, poverty relief depends on laying claim to “worker” status. Second, massive new benefits have been directed to low-income families with children and a working adult. Programs like the EITC and the Child Care and Development Fund (CCDF) provide “work supports” to the “working poor.” No longer does policy assume that “workers” can be left to live off market earnings alone. Whether old benefits are withdrawn from those who do not work or new benefits are granted to those who do, access to poverty relief now requires passing a work test in addition to a means test. That is why it matters so much that these policies do not see family caretaking as a form of work.

To end childcare invisibility, I propose that we begin accounting for care. First, we should fully incorporate childcare into the needs and resources prongs of means testing. Like food or shelter, childcare should count among the basic needs that any household with children ought to have the resources to meet. When these childcare needs are met through nonmarket care, we should count as a resource the household's savings from not having to spend income to hire third-party care, which in turn lowers its need for government assistance.

Second, accounting for care also transforms the work-testing component of antipov-erty policy. Nonmarket caretaking should satisfy work requirements when it meets childcare needs that otherwise would be met through government transfers. This conclusion follows directly from applying a standard theory of work requirements that treats work as a form of self-support that limits the need for government assistance. This theory has been thought to privilege market employment and exclude nonmarket care as nonwork, but that result no longer follows once childcare is made visible within means testing. Thus, while I build on longstanding feminist arguments for recognizing caretaking as valuable work,<sup>2</sup> my approach breaks new ground by relying on a theory of work tailored to means-tested benefits.

## I. ACCOUNTING FOR CARE WHEN ASSESSING FINANCIAL NEED

This part analyzes how childcare invisibility functions within means testing. This aspect of childcare invisibility disadvantages employed single parents relative to those who care for their own children. Among two-parent households, it disadvantages dual-earner couples relative to those with a breadwinner/caretaker division of labor; within the latter, it disadvantages the caretaker. These problems arise because means-testing systems were built on a “family wage” model of social policy, one that assumes a breadwinner/caretaker division of labor and focuses exclusively on cash transactions. Truly fixing these problems requires recognizing that childcare is needed, and provided, regardless of whether all parents are employed or whether there is an at-home caregiver; the differences concern only shifts between nonmarket and market forms of care.

### A. How Childcare Invisibility Works, and Who Gets Hurt

A means test compares what a household needs with its resources available to meet those needs—its “means.” A “standard of need” represents the minimum income on which a household can be expected to live, and from this one subtracts available income. A cash transfer makes up any difference. A longstanding concern about means-tested benefits is that they reduce the financial incentive to increase earnings. The standard economic model treats means testing as an implicit tax because when earnings rise, transfers decline. The additional earnings essentially get split between reducing government transfers and going into the earner's pockets.

If, however, increasing employment also means losing the ability to provide childcare oneself, then not only do transfers decline, but the parent also must make alternate childcare arrangements. If those arrangements cost money, then wage-earning parents can easily end up financially worse off than non-earners, despite a significantly higher cash income. This problem is especially acute for low-wage workers, because childcare costs can easily approach or exceed earnings from even full-time work. Research in the early 1990s showed that these and other “costs of working” meant that single mothers receiving public assistance could not break even financially unless they got full-time jobs paying one-and-a-half to two times the minimum wage.<sup>3</sup>

This counterintuitive result of earning more but having less arises from two related ways in which childcare is structurally invisible within conventional means-tested transfer programs. First, childcare is not included when setting the standard of need. Consequently, regardless of whether households contain only employed parents or instead include an at-home caregiver, existing policies treat them all as having the same financial needs, which do not include childcare. Second, employed parents are treated as necessarily having higher incomes available to meet these fixed needs, not accounting for how a household with an at-home caregiver obtains valuable childcare without having to pay for it in cash.

In combination, these two aspects of conventional means testing lead to overstating economic well-being and therefore understating the need for transfers. This dynamic disadvantages families that have to pay for childcare relative to those with a family caretaker. Among single-parent households, the result is a clear disadvantage for employed parents.

Similarly, among two-parent households, childcare invisibility disadvantages dual-earner couples relative to those with breadwinner/caretaker divisions of labor. The former have a higher cash income, but their higher childcare expenses remain invisible. More subtly, this result may be bad individually for caretakers in two-parent households, notwithstanding the benefit at the household level.<sup>4</sup> Holding the couple (and the breadwinner’s employment) constant, the means test conditions transfers on the caretaker remaining out of the labor market, as with a single caretaker. Additionally, holding caretaking constant, the work test conditions her access to transfers on remaining coupled to the wage-earner, for reasons described below.<sup>5</sup>

Childcare’s invisibility is no accident. To the contrary, it reflects the integration of poverty measurement and poverty relief into a broader social policy regime constructed in the early 20th century to support a family wage system.<sup>6</sup> This regime

**B. Childcare  
Invisibility as a  
Legacy of the  
Family Wage**

rests on an idealization in which citizens dwell in families headed by a married cross-sex couple, one of whom (guess which) is (at least primarily) a market wage-earner while the other (guess which) is (at least primarily) their children's caretaker.<sup>7</sup> The "breadwinner" is the "American Worker" whose market production brings resources into the household where they are consumed by his "dependents."<sup>8</sup>

The influence of family wage norms dovetails with a conception of household economic well-being grounded in cash purchases in consumer markets. Households are poor when they lack the money to buy things. This understanding of poverty reflects broader tendencies to identify economic activity with markets and to identify social citizenship with participation in a "consumer's republic."<sup>9</sup> If childcare is assumed to be provided by the mother, then childcare is excluded from household needs measured in cash. And because caretaking (and housework generally) produces no cash for use in consumer markets, it adds nothing to a cash measure of resources available to meet household needs.

Fingerprints of this consumerist family wage model lie all over the origins of current poverty measurement. When the first federal poverty thresholds were developed in the early 1960s, officials based them on surveys of actual expenditure patterns.<sup>10</sup> Childcare was virtually absent from these family budgets because employment was a relative rarity among mothers, especially married ones with young children,<sup>11</sup> and even then care almost always was provided without charge by a relative.<sup>12</sup> Since then, official poverty measures have been adjusted only for inflation, not for changes in the composition of household spending.

In this way, the work/family patterns of the 1950s literally remain built into contemporary poverty measurement. As a result, a household of three adults has the same poverty line as a household of one adult and two children, and the latter has the same poverty line regardless of whether the adult is available as a caretaker. Variations in childcare costs among these three-person households are invisible to official determinations of economic need.<sup>13</sup>

Childcare's invisibility in poverty measurement carries over into assessments of financial need that determine eligibility for antipoverty programs. For instance, TANF financial eligibility standards typically aim to establish whether a household has "sufficient income to meet their most basic needs" or the like.<sup>14</sup> Yet childcare never appears in any state's explanation of what those basic needs include,<sup>15</sup> even when they provide detailed lists including not only housing, clothing, food, utilities, and health care but also "recreation"<sup>16</sup> or even "scout uniforms."<sup>17</sup>

Fresh approaches to policy redesign open up once we trace means testing's pathological treatment of childcare to its grounding in a family wage model. The solutions lie in rejecting the assumptions both that childcare is provided outside the market and that markets are the exclusive determinant of a household's economic well-being. This section argues for taking the first of these two steps by treating childcare as a basic need of all households with children.

## C. Childcare as a Basic Need

Imagine if poverty measurement and antipoverty policy began with a household in which all adults are employed full-time. Childcare would seem indistinguishable from other necessities like food, clothing, and shelter. That is precisely my proposal: simply treat childcare as something that all households with children need.

### 1. Basic Needs Budgeting for All Families

Others have taken a step in this direction with "basic needs budgets" or "self-sufficiency standards."<sup>18</sup> These incorporate *employed* parents' childcare expenses into calculations of how much income a household needs to meet its basic needs. Of course, these expenses vary by the number, age, and health status of the household's children, and so the amount added varies accordingly. These proposals fall short, however, because they remain anchored in the cash economy of market earnings and market spending.<sup>19</sup> Childcare becomes visible only when all adults are in the labor market or otherwise systematically unavailable to provide care. With a nonmarket caretaker present, childcare again disappears from the analysis.

This market-centered approach to childcare mirrors an early poverty measurement approach to subsistence food production by farm families. Because farm families purchased less food, they were deemed to have lower household needs, reflected in lower poverty thresholds that persisted until 1981.<sup>20</sup> By analogy to childcare, one might treat food purchases as a "work expense." After all, only when farmers stop farming and enter the labor market do they need to buy groceries for dinner, and so only then is food included in the tally of household needs. By this logic, families that farm do not need to eat, just like families that include a nonmarket caretaker do not need to ensure their kids are cared for. This is silly. What varies in these cases is not a household's economic needs, but only whether those needs are met through nonmarket production or market consumption.<sup>21</sup> Accordingly, I would extend the basic needs budgeting approach by incorporating childcare into the standard of need for all households, not only those that lack a nonmarket caretaker. Whether or not you grow your own, we all need to eat.

Alongside work-based welfare reform, a consensus emerged that lack of childcare could justify a transfer recipient's failure to work and that, to avoid this situation, public resources should be devoted to providing access to childcare. The precise contours of these principles remain controversial,<sup>22</sup> but even the Heritage

### 2. Against the "Work Expense" Model



Foundation concedes that “if low-skilled single mothers are moved into the labor force, childcare assistance must be provided.”<sup>23</sup> Funding for childcare assistance to low-income households has expanded massively in tandem with the growing emphasis on work.<sup>24</sup>

This work-expense framework yields two kinds of institutional responses. The first deducts some or all childcare expenses from gross earnings; the resulting net income becomes the measure of household resources for the purpose of poverty measurement or means testing.<sup>25</sup> The second, now dominant, approach leaves means-tested cash transfers untouched but creates a separate childcare benefit tied to employment. Rather than providing additional cash to cover childcare expenses, direct childcare assistance relieves parents of out-of-pocket expenses.

Both approaches shift childcare costs from employed parents to the state. But if not to meet basic needs, then why? The usual rationale is some variant on “supporting work” or enabling transfer recipients to comply with work requirements. These rationales lose track of *why* work is promoted within a means-tested transfer system. Instead, antipoverty policy slides toward making employment an end in itself.<sup>26</sup>

The simple problem is that childcare assistance is enormously expensive. For a household with a toddler and a school-age child, a typical state CCDF program would authorize childcare subsidies totaling about \$11,500 per year.<sup>27</sup> This roughly equals the median maximum household benefit that a household of three without any income would get from TANF and SNAP combined.<sup>28</sup> It also roughly equals full-time, full-year earnings at the minimum wage, net of federal payroll taxes.<sup>29</sup>

Linking childcare assistance to employment collides with the theory that employment reduces reliance on transfers. What is the point of these childcare expenses? To support work. And why is work important? Because it reduces transfers in favor of self-support.<sup>30</sup> Except that work may not reduce transfers once childcare becomes visible and work triggers new transfers. This contradiction is masked by keeping a separate set of books for cash assistance (which employment reduces) and childcare assistance (which it increases).<sup>31</sup>

#### D. How Means-Testing Can Account for Care

Treating childcare as a basic need alters both components of means testing: setting a standard of need and measuring the resources available to meet those needs. This section sketches how both components can account for care.

##### 1. Setting the Standard of Need

Treating childcare as a basic need is straightforward with regard to the household needs side of means testing. Instead of one cash assistance program that ignores

childcare and a separate childcare assistance program for some employed parents, there should be a single program with a comprehensive standard of need, childcare included. In practical terms, this means integrating what are now separate systems for providing income support through TANF and childcare assistance through CCDF.

Integrating cash and childcare assistance can be accomplished by taking existing standards of need and adding to them the reimbursement rates already authorized for childcare assistance. Schematically, if existing cash assistance programs assume that a household needs \$10,000 per year, exclusive of childcare, and existing childcare assistance programs authorize government expenditures of \$6,000 per year to provide care for a low-income worker's child, then the two can be combined to yield a \$16,000 standard of need. Existing reimbursement rates vary with children's age, special needs, and types of care, and so too would the new standards of need. They would also reflect how needs vary with the number of children. In the example above, a second child would raise the standard of need to \$22,000. Of course, there's nothing magical about existing standards of need and reimbursement rates. They could be too high or too low, but those concerns are distinct from the question of integration.

The next step addresses the household resources side of means testing. Here, the challenge arises from implementing the proposition that childcare does not vanish as a basic need when a family caretaker is present. Instead, that need is met by the caretaker directly, not through purchasing care as a consumer service. Correspondingly, the means-testing calculus must widen its net beyond market transactions in order to capture the household's nonmarket resources for meeting its needs.

## 2. Measuring Available Resources

Failing to account for this nonmarket resource would generate a windfall in transfers to households that include a nonmarket caretaker. For those households, measuring resources in cash alone would generate a transfer large enough to pay for childcare *and* other needs. However, with a caretaker present who spends her time, but not her money, on childcare, the portion of the transfer designated for childcare ends up left over, even after all other basic needs are paid for. Such windfalls arise whenever resources that the household can use to satisfy its needs are not counted in the means test's measurement of income.

That windfalls result from unmeasured income provides a clue toward preventing them. Consider the more familiar example of food. Like \$200 in cash, a \$200 gift certificate to a grocery store should count as "income" because it substitutes for \$200 in transfers otherwise needed to purchase food.<sup>32</sup> Moving further from cash

toward in-kind resources, the same analysis applies to a gifted shopping cart full of \$200 in essential groceries. Not counting any of these as income would create a \$200 windfall relative to what the transfer is designed to provide.

Windfalls to nonmarket caretakers can be avoided using the same basic techniques that capture in-kind income. Like a full shopping cart and food expenses, nonmarket caretaking allows the household to avoid childcare expenses. Therefore, the solution is to credit the household with “imputed income,” an accounting device that treats avoided expenses as equivalent to income used to pay the expense.<sup>33</sup> In this case, the amount of imputed income from caretaking would equal the amount designated for childcare in the standard of need.

By anchoring imputed income in the standard of need, my approach aims to capture realistically the trade-off between devoting time (but not money) to childcare versus devoting time to employment and its earnings to childcare. For a household with one sixteen-year-old who has no special needs, caretaking would yield no imputed income because existing childcare programs would not fund any care at all (thirteen is the usual age cutoff)—reflecting the judgment that adult supervision outside school hours is not necessary, even if it may be desirable. In contrast, the imputed income from caring for three pre-schoolers might well *exceed* the parent’s opportunity costs in the labor market. It all depends on the number, age, and health status of the children, and the price and quality of acceptable childcare. Generally speaking, for low-wage workers, wage income will be roughly equivalent to imputed income from caring for two preschool-age children without special needs.

## II. ACCOUNTING FOR CARE WHEN ASSESSING WORK

Just as traditional means testing focuses on spending cash in consumer markets, traditional work requirements focus on earning cash in labor markets. Likewise, incorporating nonmarket income and consumption into means testing provides the foundation for incorporating nonmarket production into work requirements.

### A. The Invisible Work of Family Caretakers

Exactly what activities satisfy current work requirements in TANF and related programs is quite a bit messier than often assumed.<sup>34</sup> Nonetheless, paid employment clearly provides the paradigm. It is equally clear that policymakers have invoked “work” to justify denying transfers to family caretakers. With only the narrowest and most tenuous of exceptions, caring for a member of one’s own household never counts as work in contemporary welfare policy.

In its raw form, the charge against caretakers is that they are sitting at home doing “nothing.” Thus, the family caretaker is no different than the proverbial Malibu

surfer who looms large in philosophical debates over linking redistribution to work.<sup>35</sup> This “nothing” charge dovetails with childcare’s economic invisibility in the machinery of means testing and poverty measurement. A cash accounting system attributes zero income to the family caretaker, in contrast to the thousands that market employment would generate.

Assessing the “doing nothing” charge requires a theory of work requirements.<sup>36</sup> Many virtues are attributed to employment, but which provide the “something” that makes it work and not “nothing”? This section offers an answer distilled from leading arguments for work requirements and structural features of actual policies.

**B. Defining Work  
As Meeting Basic  
Needs**

I begin with a simple, familiar rationale for means-tested transfers: society owes support to people who otherwise could not meet their basic needs through no fault of their own. Faultlessness includes having made reasonable efforts at self-support through one’s own work, where possible.<sup>37</sup>

An expectation of reasonable self-support integrates means testing and work testing at both theoretical and practical levels. A means test identifies those in need by comparing resources to some threshold. Work requirements capture the idea that individuals also can “spend” their time on activities that help meet their economic needs. Someone who could work but does not is someone who has the means to satisfy his basic needs, notwithstanding low income.

This connection between work and need operates automatically within a system of cash accounting. When need is reduced to cash purchasing, work can be reduced to generating cash income.

As noted above, however, means tests have always sought to capture at least some noncash resources. And when in-kind receipts should count as income, the activity producing them should count as work. If an employee gets paid in free rent rather than cash, she still has earnings and still works,<sup>38</sup> though the issue rarely arises because minimum wage law generally mandates cash payment. This analysis applies with equal strength when a household member produces goods or services that are consumed within the household. There is no functional difference between employment paid with a \$200 grocery store gift certificate and subsistence farming that produces \$200 worth of food.

This approach to identifying nonmarket work draws on, but is narrower than, techniques for identifying nonmarket work based on its substitution for market production. In the means-testing context, what counts is not all economic production but only production of resources that the worker’s household can use to meet its

basic needs. A volunteer firefighter substitutes for a paid firefighter, but this does not put food on her table. Similarly, walking one's own dog substitutes for hiring a dog walker, but these economic benefits to the household do not concern basic household needs. Accordingly, unlike the farmer, neither the volunteer nor the dog walker would satisfy work requirements on means-tested transfers, even though all three are in some sense engaged in nonmarket work that substitutes for market production.<sup>39</sup>

### C. How Work Requirements Can Account for Care as Self-Support

Tying work status to self-support offers an important variation on more familiar feminist theories of nonmarket care work as a basis for transfer receipt. These theories generally rely on the point that nonmarket care is *socially* valuable, even though the caretaker does not get paid.<sup>40</sup> Producing a social contribution triggers a claim to share in the fruits of this labor (or to relief from the burdens of producing them).<sup>41</sup> On that analysis, the resulting transfer is analogous to a worker's *wage*, a share of what she produced. For this reason, social contribution theories typically point toward economic claims by all caretakers, regardless of their poverty status.<sup>42</sup>

In contrast, my argument is that low-income caretakers are members of the working poor: they lack the resources to meet basic needs despite reasonable efforts at self-support. The caretaker covers childcare directly but then needs cash transfers with which to buy other necessities. The low-wage worker generates cash to purchase those same necessities, but at the cost of needing yet more cash to obtain childcare. Either way, "work alone is not enough,"<sup>43</sup> and the government steps in to make up the difference with transfers. Such transfers to a low-wage employee do not compensate her for the value of her work, nor do they do so for the low-income caretaker.

Self-support distinguishes childcare from other unpaid but productive work that contributes to society but does not substitute for transfers to the worker,<sup>44</sup> and it avoids the need to interrogate either the magnitude of these contributions or the sufficiency of any compensating private gains.<sup>45</sup> In a delightful twist on gender stereotypes, the caretaker is best compared not to the selfless volunteer but to the yeoman farmer.

### How Much Work?

Programs that support the working poor provide transfers when work falls short of satisfying all household needs. Current policies accept full-time, minimum wage employment as a sufficient effort at self-support. They do not insist on sixty-hour weeks or obtaining a \$12 hourly wage. For instance, TANF states work requirements in weekly hours, typically in the thirty- to forty-hour range.

Nonmarket caretaking presents obvious challenges to this temporal accounting system, but the methods to tackle them have been developed to address similar difficulties with self-employment. Like caretaking, self-employment lacks the hierarchically imposed time discipline and separation of “working time” from “personal time” characteristic of employment.<sup>46</sup> To overcome these challenges, TANF rules permit imputation of time from income. These rules attribute to the self-employed individual hours of work equal to their net business income divided by the minimum wage.<sup>47</sup> This method makes it unnecessary to disentangle various uses of time, so long as the transfer recipient gets the job done, that is, earns the money that offsets transfers.

Applying these methods, nonmarket caretaking can be integrated into hours-based work requirements. Hours worked would equal imputed income divided by the minimum wage. If caretaking substitutes for \$12,000 in annual transfers to cover childcare expenses, it is equivalent to full-time work. But if it substitutes for only \$6,000, the caretaker would receive credit only for half-time work. In this way, the techniques for quantifying income carry over to attributing hours of work. As a result, caretaking satisfies work requirements to the same extent as a minimum wage job with an equal contribution to self-support.

Accounting for care within work requirements has some counterintuitive effects when applied to two-parent families. In the simple case where both parents are required to work, the effects track those for single-parent households. Nonmarket caretaking shifts from being a basis for disqualification by work requirements to one acceptable method of satisfying them.

*Two-Parent  
Couples*

What complicates matters is that currently work requirements generally do not apply to both parents. Instead, both TANF and the EITC primarily utilize what I have labeled elsewhere a “breadwinner priority” structure.<sup>48</sup> They require only one adult per household to work. Contrary to the widespread interpretation that welfare reform required low-income mothers to substitute employment for caretaking, instead it has required them to be part of “working families” and permitted a breadwinner/caretaker division of labor. My approach continues to permit caretaking within a two-parent family, but it treats caretakers as satisfying work requirements rather than exempt from them.<sup>49</sup>

Treating caretaking as work provides a basis for reconciling aspects of the existing breadwinner priority policy with a self-support theory of work requirements. Allowing a married parent to be a caretaker and receive need-based transfers is difficult to justify if caretaking is not work when single parents do it. A self-support

analysis also implies that sometimes both parents should be expected to enter the labor market, something the current breadwinner priority model never does. If a family's only child is sixteen, the state will not provide childcare assistance even if both parents have jobs; therefore, staying home as a caretaker does not further self-support. Allowing it nonetheless suggests an inappropriate thumb on the scale in favor of a (likely gendered) breadwinner/caretaker division of labor.

**CONCLUSION:  
WHAT'S LEFT  
OF WORK  
REQUIREMENTS?**

**M**y proposal obviously calls for a substantial change in the way work requirements currently are structured in major programs designed to assist low-income parents. Given the political and academic popularity of those work policies,<sup>50</sup> skepticism, even scorn, is predictable and understandable. That granted, my approach has three features that differentiate it from a simple recycling of old debates about work and care, features that could create some space for new thinking. First, it integrates caretaking into work requirements; it does not reject work requirements or propose exemptions from them.<sup>51</sup> Second, it treats caretaking as work for reasons that are theoretically and practically intertwined with supporting childcare assistance for employed parents.<sup>52</sup> Third, it integrates the treatment of work and care across single- and two-parent households, creating the potential to leverage policies and politics that already support caretaking in the latter.<sup>53</sup>

My approach is grounded in the deep support for childcare assistance, an area where costly *expansions* in need-based transfers have remained politically popular.<sup>54</sup> So, too, is the idea that parents should get to control basic child-rearing decisions. Combining the two creates an opening for allowing low-income parents to choose between getting childcare assistance while they earn money and getting money while they provide childcare.<sup>55</sup>

Progressives too often have ceded the discourse of work, responsibility, and self-reliance to the right. The arguments offered here may hold out some hope that we can embrace the cause of the working poor yet fend off its evil twin, abandonment of those who work outside conventional labor markets.







- \* This essay is an abridged version of Noah D. Zatz's article, *Supporting Workers By Accounting for Care*, 5 HARVARD LAW & POLICY REVIEW 45 (2011).
- \*\* This paper derives from one written and presented while I was a fellow at the Program in Law and Public Affairs at Princeton University, and I gratefully acknowledge the material and intellectual support LAPA and its community provided. I also have benefitted greatly from comments on related work from Robin Lenhardt and Viviana Zelizer and from workshop participants at UCLA School of Law; the Washington University, St. Louis, workshop on Work, Family, and Public Policy; and the 2009 Emerging Family Law Scholars & Teachers Conference, as well as from the editors of the Harvard Law & Policy Review. Invaluable research assistance was provided by Elizabeth Adams, Kelly Hoehn, and Melissa Rose.
1. NANCY FOLBRE, VALUING CHILDREN: RETHINKING THE ECONOMICS OF THE FAMILY 45–64 (2008).
  2. See, e.g., MARTHA ALBERTSON FINEMAN, THE NEUTERED MOTHER, THE SEXUAL FAMILY AND OTHER TWENTIETH CENTURY TRAGEDIES (1995); Sylvia A. Law, *Women, Work, Welfare, and the Preservation of Patriarchy*, 131 U. PA. L. REV. 1249 (1983).
  3. See KATHRYN EDIN & LAURA LEIN, MAKING ENDS MEET: HOW SINGLE MOTHERS SURVIVE WELFARE AND LOW-WAGE WORK 81, 230 (1997).
  4. Noah D. Zatz, *Revisiting the Class Parity Analysis of Welfare Work Requirements*, 83 SOC. SERV. REV. 313, 316 (2009).
  5. See discussion *infra* at note 48 and surrounding text.
  6. See generally LINDA GORDON, PITIED BUT NOT ENTITLED: SINGLE MOTHERS AND THE HISTORY OF WELFARE, 1890-1935 (1994); ALICE KESSLER-HARRIS, IN PURSUIT OF EQUITY: WOMEN, MEN, AND THE QUEST FOR ECONOMIC CITIZENSHIP IN 20TH-CENTURY AMERICA (2001); THEDA SKOCPOL, PROTECTING SOLDIERS AND MOTHERS: THE POLITICAL ORIGINS OF SOCIAL POLICY IN THE UNITED STATES (1992).
  7. Ann Shola Orloff, *Gender and the Social Rights of Citizenship: The Comparative Analysis of Gender Relations and Welfare States*, 58 AM. SOC. REV. 303, 308 (1993). On the racialization of the family wage system, see GORDON, *supra* note 6, at 4, 7, 293–94; SUZANNE METTLER, DIVIDING CITIZENS: GENDER AND FEDERALISM IN NEW DEAL PUBLIC POLICY 172 (1998); Dorothy E. Roberts, *The Value of Black Mothers' Work*, 26 CONN. L. REV. 871, 875 (1994).
  8. See generally Nancy Fraser & Linda Gordon, *A Genealogy of Dependency: Tracing a Keyword of the U.S. Welfare State*, 19 SIGNS 309 (1994).
  9. See FELICIA KORNBLOH, THE BATTLE FOR WELFARE RIGHTS 40 (2007); see also LIZABETH COHEN, A CONSUMERS' REPUBLIC: THE POLITICS OF MASS CONSUMPTION IN POSTWAR AMERICA (2003).
  10. See Gordon M. Fisher, *The Development and History of the Poverty Thresholds*, 55 SOC. SECURITY BULL., no. 4, Winter 1992 at 3, 4, available at <http://www.ssa.gov/policy/docs/ssb/v55n4/v55n4p3.pdf>; Mollie Orshansky, *Counting the Poor: Another Look at the Poverty Profile*, 28 SOC. SECURITY BULL., no. 1, Jan. 1965 at 3, 8, available at <http://www.ssa.gov/policy/docs/ssb/v28n1/v28n1p3.pdf>.

11. BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, LABOR FORCE STATISTICS DERIVED FROM THE CURRENT POPULATION SURVEY: BULLETIN 2307, 801, tbl. C-12 (1988).
12. SONYA MICHEL, CHILDREN'S INTERESTS/MOTHERS' RIGHTS: THE SHAPING OF AMERICA'S CHILD CARE POLICY 64 (1999).
13. See FOLBRE, *supra* note 1, at 45, 52.
14. ME. REV. STAT. ANN. tit. 22, § 3769-C (2004 & Supp. 2011). For the formulation "reasonable subsistence compatible with decency and health," see ALA. CODE § 38-4-1 (Lexis-Nexis 1992); KAN. STAT. ANN. § 39-709 (2000 & Supp. 2010); MISS. CODE ANN. § 43-17-5 (Supp. 2011); TEX. HUM. RESOURCES CODE ANN. § 31.003 (West 2001); WYO. STAT. ANN. § 42-2-103 (2011).
15. This statement is based on my review of the statutes and regulations governing the TANF programs in all states and the District of Columbia, as they stood on Aug. 1, 2008.
16. CAL. WELF. & INST. CODE § 11452 (West 2001).
17. CONN. DEP'T OF SOC. SERVS., UNIFORM POLICY MANUAL § 4510.05, available at <http://www.sharinglaw.net/elder/UPM.htm> (last updated Oct. 26, 2004).
18. See, e.g., Diana M. Pearce, *The Self-Sufficiency Standard: A New Tool for Evaluating Anti-Poverty Policy*, 10 POVERTY & RACE, no. 2, Apr. 30, 2001 at 3; Trudi J. Renwick & Barbara R. Bergmann, *A Budget-Based Definition of Poverty With an Application to Single-Parent Families*, 28 J. HUM. RESOURCES 1 (1993).
19. FOLBRE, *supra* note 1, at 57-59.
20. THOMAS A. CARLIN ET AL., U.S. DEP'T OF AGRIC., THE FARM DIFFERENTIATION IN THE POVERTY THRESHOLD: SHOULD IT BE CHANGED?, in ASPECTS OF WELFARE AND POVERTY IN RURAL AMERICA: TWO ISSUE BRIEFS 11 (1979), available at <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED187510>.
21. Noah D. Zatz, *What Welfare Requires From Work*, 54 UCLA L. REV. 373, 398-400 (2006).
22. LAWRENCE M. MEAD, THE NEW POLITICS OF POVERTY: THE NONWORKING POOR IN AMERICA 121-22 (1992).
23. Brian Riedl, *The Myth of a Child Care Crisis*, THE HERITAGE FOUND. (May 2, 2003), <http://www.heritage.org/Research/Reports/2003/05/The-Myth-of-a-Child-Care-Crisis> (on file with the Harvard law School Library).
24. STAFF OF H. COMM. ON WAYS & MEANS, 110TH CONG., 2008 GREEN BOOK: BACKGROUND MATERIAL AND DATA ON THE PROGRAMS WITHIN THE JURISDICTION OF THE COMMITTEE ON WAYS AND MEANS, at 9-2 to 9-3, 9-11 (Comm. Print 2008), available at <http://democrats.waysandmeans.house.gov/singlepages.aspx?NewsID=10490>.
25. MEASURING POVERTY: A NEW APPROACH 207-10 (Constantine F. Citro & Robert T. Michael eds., 1995); Jonathan Zasloff, *Children, Families, and Bureaucrats: A Prehistory of Welfare Reform*, 14 J.L. & POL. 225, 235-36 (1998) (describing "child care disregard").
26. Zatz, *supra* note 21, at 404-05, 424-25.
27. Author's calculation of population-weighted average maximum state CCDF reimbursement rates for family day care, based on CHILD CARE BUREAU, U.S. DEP'T OF HEALTH

& HUM. SERVS., CHILD CARE AND DEVELOPMENT FUND: REPORT OF STATE AND TERRITORY PLANS, FY 2010–2011, at 15, tbl. 6 (2010), *available at* <http://www.acf.hhs.gov/programs/occ/ccdf/plans/sp2010-2011.pdf>. For an infant and a toddler, the figure would be almost \$14,000.

28. This benefit would total \$11,350 per year. Author's calculation, based on LIZ SCHOTT & IFE FINCH, CTR. ON BUDGET & POL'Y PRIORITIES, TANF BENEFITS ARE LOW AND HAVE NOT KEPT PACE WITH INFLATION app.4 (2010), *available at* <http://www.cbpp.org/files/10-14-totanf.pdf>. Note that a full-time, full-year minimum wage worker would still receive some of these benefits, though at a much reduced level, but would also be eligible for the maximum EITC in the vicinity of \$5,000. Integration with the EITC is theoretically attractive but raises additional issues. *Cf.* Nancy C. Staudt, *Taxing Housework*, 84 GEO. L.J. 1571 (1996) (arguing for inclusion of imputed income from household labor within the income tax base and considering interactions with the EITC).
29. The full year earnings would be calculated as follows: (50 weeks/year \* 35 hours/week \* \$7.25/hour) \* (100% - 7.65% FICA) = \$11,700.
30. See discussion *infra* Part II.C.
31. Extending the time frame could resolve the contradiction in conjunction with rapid earnings growth, but the available evidence is discouraging. See Zatz, *supra* note 21, at 405–18.
32. Zatz, *supra* note 21, at 395–400; see also James Tobin et al., *Is a Negative Income Tax Practical?*, 77 YALE L.J. 1, 12 (1967). In some contexts certain in-kind receipts are excluded from income calculations because of administrative difficulties in valuation, see, for example, *Maisonet v. N.J. Dep't of Hum. Servs.*, 643 A.2d 1038 (N.J. Super. Ct. App. Div. 1994), but these administrative exceptions illustrate the underlying rule.
33. See Staudt, *supra* note 28, at 1576 n.18; KATHARINE G. ABRAHAM & CHRISTOPHER D. MACKIE, BEYOND THE MARKET: DESIGNING NONMARKET ACCOUNTS FOR THE UNITED STATES 56 (2005); FOLBRE, *supra* note 1, at 122–24.
34. See Noah Zatz, *Welfare to What?*, 57 HASTINGS L.J. 1131, 1184 (2006); Zatz, *supra* note 21, at 384.
35. See PHILIPPE VAN PARIJS, REAL FREEDOM FOR ALL 89 (1995); Kirk J. Stark, *Enslaving the Beachcomber: Some Thoughts on the Liberty Objections to Endowment Taxation*, 18 CAN. J.L. & JURIS. 47 (2005).
36. See generally Zatz, *supra* note 21.
37. See Amy L. Wax, *Disability, Reciprocity, and "Real Efficiency": A Unified Approach*, 44 WM. & MARY L. REV. 1421, 1446 (2003).
38. Indeed, in 1973, Congress amended the Food Stamps statute specifically to count employer-provided housing as income, thereby "equalizing the food stamp benefits of a person who receives his earnings entirely in the form of wages and a person who actually has the same income but receives part of it in kind rather than in cash." *Anderson v. Butz*, 428 F. Supp. 245, 253 (E.D. Cal. 1975).
39. The same basic point applies to delimiting the relevant sources of imputed income

for the purpose of measuring resources in the means test. A standard objection to imputed income techniques is that, once one begins with specific cases like childcare, there is no principled basis not to attribute “income” to every practice valued by the person doing it: if walking the dog substitutes for hiring a pet sitter or playing solitaire substitutes for buying theater tickets, why not impute income to each at the value of their market substitutes? See Thomas Chancellor, *Imputed Income and the Ideal Income Tax*, 67 OR. L. REV. 561, 561–62 (1988). The means-testing context provides an answer: childcare is included in the basic needs budget but pet sitting and theater tickets are not.

40. See, e.g., FINEMAN, *supra* note 2, at 9; Nancy Folbre, *Children as Public Goods*, 84 AM. ECON. REV. 86, 87–88 (1994). My argument is more akin to Staudt’s argument for taxing housework based on its contribution to household consumption, see Staudt, *supra* note 28, at 1618–22, 1627–29, though she shies away from fully equalizing nonmarket and wage income for the purpose of access to redistributive transfers like the EITC, see *id.* 1637–38.
41. See, e.g., ANNE L. ALSTOTT, NO EXIT: WHAT PARENTS OWE THEIR CHILDREN AND WHAT SOCIETY OWES PARENTS 58–61 (2004); FOLBRE, *supra* note 1, at 178–83.
42. See ALSTOTT, *supra* note 41, at 75–76; EVA FEDER KITTAY, LOVE’S LABOR: ESSAYS ON WOMEN, EQUALITY, AND DEPENDENCY 142–43 (1999); GWENDOLYN MINK, WELFARE’S END 135 (1998).
43. Robert Haveman, *When Work Alone Is Not Enough*, in ONE PERCENT FOR THE KIDS 40 (Isabel V. Sawhill ed., 2003).
44. For arguments equating nonmarket care and volunteering, see Stuart White, *Fair Reciprocity and Basic Income*, in REAL LIBERTARIANISM ASSESSED: POLITICAL THEORY AFTER VAN PARIJS 136 (Andrew Reeve & Andrew Williams eds., 2003); Iris Marion Young, *Autonomy, Welfare Reform, and Meaningful Work*, in THE SUBJECT OF CARE: FEMINIST PERSPECTIVES ON DEPENDENCY 40 (Eva Feder Kittay & Ellen K. Feder eds., 2003). Different lines of argument may justify volunteering as a form of job training or a demonstration of willingness to work.
45. See, e.g., Mary Anne Case, *How High the Apple Pie? A Few Troubling Questions About Where, Why, and How the Burden of Care for Children Should Be Shifted*, 76 CHI.-KENT L. REV. 1753, 1775 (2001); Katherine M. Franke, *Theorizing Yes: An Essay on Feminism, Law, and Desire*, 101 COLUM. L. REV. 181, 191–92 (2001).
46. Zatz, *supra* note 21, at 431.
47. 45 C.F.R. § 261.60(d) (2011). Similarly, the EITC uses earnings as a proxy for work effort by maximizing transfers at earnings levels approximating those from full-time, minimum wage employment.
48. Zatz, *supra* note 4, at 321. This predominates in state TANF rules, too, but some states require both adults to work.
49. The two-parent context raises difficulties in attributing carework to individual parents. ALSTOTT, *supra* note 41, at 181–84; Staudt, *supra* note 28, at 1623–24. However, because means-tested programs generally calculate and pay benefits on a household

basis, one could require eighty hours of work in the aggregate, not necessarily forty by each parent. Similarly, we could attribute hours of care (and its imputed income) based on eighty hours minus the greater of forty hours or both parents' aggregate time in waged work.

50. See, e.g., Ron Haskins, *Liberal and Conservative Influences on the Welfare Reform Legislation of 1996*, in *FOR BETTER AND FOR WORSE: WELFARE REFORM AND THE WELL-BEING OF CHILDREN AND FAMILIES* 9 (Greg J. Duncan & P. Lindsay Chase-Lansdale eds., 2001); Hugh Heclo, *The Politics of Welfare Reform*, in *THE NEW WORLD OF WELFARE* 169 (Rebecca M. Blank & Ron Haskins eds., 2001).
51. For two examples of arguments against work requirements, based in part on their exclusion of caretaking, see Carole Pateman, *The Patriarchal Welfare State*, in *DEMOCRACY AND THE WELFARE STATE* 231 (Amy Gutmann ed., 1988); Anne L. Alstott, *Work vs. Freedom: A Liberal Challenge to Employment Subsidies*, 108 *YALE L.J.* 967 (1999).
52. Cf. ALSTOTT, *supra* note 41 (proposing caretaker resource accounts that would benefit both employed parents and family caretakers).
53. See Zatz, *supra* note 4, at 321.
54. Sandra K. Danziger, *The Decline of Cash Welfare and Implications for Social Policy and Poverty*, 36 *ANN. REV. SOC.* 523, 524, 537–38 (2010).
55. Cf. Press Release, Hillary Clinton for President, *Hillary Clinton's Agenda for Working Families: Helping Parents Balance Work & Family* (Oct. 16, 2007) (proposing “to allow qualifying, low-income parents who want to stay at home rather than place their children in child care to receive child care subsidies through the Child Care Development Block Grant”); NAT'L P'SHIP FOR WOMEN & FAMILIES, *AT-HOME INFANT CARE (AHIC): A SIDE-BY-SIDE COMPARISON OF FEDERAL AND STATE INITIATIVES* (2005), available at <http://www.nationalpartnership.org/site/DocServer/AHICchartOct05.pdf?docID=1048>; Caregivers and Veterans Omnibus Health Services Act of 2010 (“Act”), 38 U.S.C. §§1720G(a)(3)(C)(ii) (Supp. IV 2010) (making family caregivers of injured veterans eligible to receive a cash stipend “not less than the monthly amount a commercial home health care entity would pay an individual . . . to provide equivalent personal care services to the eligible veteran”).

