LIBERTY, PROPERTY, AND THE LAW

Contract - Freedom and Restraint

Edited by
Richard A. Epstein

Liberty, Property, and the Law

A Collection of Essays

Series Editor

Richard A. Epstein

University of Chicago Law School

Series Contents

Contract — Freedom and Restraint

Edited with introductions by

Richard A. Epstein
University of Chicago Law School

Routledge
Taylor & Francis Group
New York London

First published by Garland Publishing, Inc.

This edition published 2011 by Routledge:

Routledge
Taylor & Francis Group
711 Third Avenue
New York, NY 10017

Routledge
Taylor & Francis Group
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN

Library of Congress Cataloging-in-Publication Data

Contract — freedom and restraint / edited with introductions by Richard A. Epstein
p. cm. — (Liberty, property, and the law ; 4)
Includes bibliographical references.
ISBN 0-8153-3558-X (alk. paper) — ISBN 0-8153-3543-1 (set)
1. Liberty of contract—United States. 2. Contracts—United States.
I. Epstein, Richard Allen, 1943– II. Series.

KF807 .C66 2000
346.7302'2—dc21 00-024524

POD ISBN: 9780415532563
Set ISBN: 9780815335436
Vol 1: 9780815335559
Vol 2: 9780815335566
Vol 3: 9780815335573
Vol 4: 9780815335580
Vol 5: 9780815335597

Contents

This page intentionally left blank

Series Introduction

A Universal Topic

The materials in this collection are drawn from many disciplines, including economics, law, philosophy and political science. Yet they are all directed to a topic that is worthy of examination from multiple perspectives: "Liberty, Property and the Law." Stated in this general form, this topic is as broad as law itself. The relationship of liberty and property to the law surfaces whenever and wherever people interact with each other under the command and control of the sovereign. Those who hold sovereign power may choose to protect liberty and property or to undermine it. But the regrettably high frequency of political abuse throughout the world does not justify the exercise of arbitrary legal power; nor does it limit human aspirations for a sound legal and social order to block political excesses.

As we reflect on these matters, the normative element of legal inquiry shines through the clouds of deviant social practice. Virtually all legal systems in the English (or common law) tradition — or, for that matter, in the Roman (or civil law) tradition — start from the central proposition that each person is entitled to exclusive control of his or her person and property, free from invasions by other individuals. The Lockean theory of social contract regards the state's purpose as protecting the "lives, liberties and estates" of individuals, or what we would call today their life, liberty and property. The United States Constitution, like many others that are modeled on it, explicitly provides that "no person shall be deprived of life, liberty or property without due process of law." Even if "due process" is given a narrow procedural interpretation — and it frequently is not — modern constitutions often supply substantive protection to property with something akin to the takings clause of the United States Constitution: "Nor shall private property be taken for public use, without just compensation."

These three (and other) points of reference show that the ideas of liberty and property are, in some sense, inseparable from the Western legal tradition — one must be careful in the choice of words — of limited and democratic government. To set the selections in these five volumes into their larger context, it is useful to outline the classical liberal synthesis that revolves around liberty and property, and then indicate the vulnerable points that expose it to counterattack. As editor of this collection, I make no secret of my defense of classical liberalism against both its traditional and modern

critics. The task on this occasion, however, is not to resolve the profound issues in this dispute, but to set out the central arguments on both sides of the debate to help readers decide for themselves.

The Classical Liberal Synthesis

The classical legal tradition begins with an account of human nature that seeks to explain how people behave under pervasive and inevitable conditions of scarcity. Whether it looks to biological or societal forces, the classical tradition emphasizes the self-interest of human beings whose aggressive tendencies make law necessary for the creation and preservation of social order. In its extreme (Hobbesian) version, self-interest is the only motive for individual human actors. In its more moderate version, this self-interest is tempered by the bonds of natural love and affection within families and towards friends, and perhaps by a sense, as Hume put it, of "confin'd generosity" to other human beings.

Starting from this account of human nature, the role of liberty and property can be defended both on natural law and consequentialist, typically utilitarian, theories. Although natural law and consequentialism are often set in opposition today, earlier writers tended to think of them as a cohesive whole: human beings can only flourish if governed by rules that take into account the good and bad features of human nature. More recently, however, natural lawyers have become more anti-consequentialist and less instrumentalist, fearing the indeterminate factual disputations in efforts to show how legal rules improve social utility or social wealth. They tend to treat the central postulates in defense of liberty and property as self-evident truths that can only be denied by people blind to philosophical argument and social practice. Yet on the consequentialist side, modern natural law theories are often derided as metaphysical nonsense that drives law away from the study of human and social behavior that explains and legitimizes its commands. Consequentialists rely instead on a utilitarian defense of liberty and property that promotes the long-run average advantage to all citizens, notwithstanding the occasional injustice that sound general rules may produce in a small fraction of cases. In practice, however, it is easy to identify a convergence between the two schools in support of individual liberty and property rights, despite the manifest tension in their philosophical orientations. Any disagreements between the natural lawyers and the consequentialists are tiny in comparison to the differences between capitalists and socialists. The former disagreement is between different varieties of classical liberals. The second disagreement goes to fundamental world view.

The classical liberal system, then, can be defended from two directions, but what are its central rules? The first principle is individual autonomy. To the natural lawyer, autonomy is the only principle that respects the need for self-realization of natural talents and abilities and thus gives the individual the right to control those talents and abilities. Denying individuals their personal autonomy turns them into instruments that are subject to the will and whim of another. Though it may be difficult to explain why individuals should have exclusive control over their bodies, talents, and abilities when these seem arbitrarily distributed, it is even more difficult to explain why anyone should be required to yield this control to other people. The consequentialist reaches the same basic conclusion, often by emphasizing how individual self-ownership

establishes the clear rights necessary for the emergence of a vibrant market economy. Since each person is the best judge of his or her preferences, a system that respects autonomy is likely to result in cooperative social outcomes with many benefits and few offsetting harms.

To some, autonomy sounds like a defensive concept that keeps other people at bay. But its notion of individual liberty allows people to do what they wish so long as they do not use force or fraud to interfere with the like rights of others. The internal consistency of this system appeals to the deductive instincts of the natural lawyer, while its functional advantages appeal to the practical instincts of the consequentialist. Its rights and duties keep their basic contours no matter how many people live within a society, and regardless of their wealth. The durability and versatility of the system thus operate in a timeless fashion across widely different cultures and natural conditions. From this observation, the more determined consequentialist would conclude that allowing people freedom of action within the boundaries of force and fraud presumptively maximizes overall human welfare. After all, the mere decision to undertake a single action is for the benefit of at least one person, and should prevail unless it can be shown to cause greater inconvenience to others.

One expression of individual liberty is the acquisition of property, which, in the state of nature, initially belongs to no one. Under the classical synthesis, the rule of first possession is the sole means for moving resources from the state of nature to private ownership, all without government intervention. A system of private property establishes boundaries between individuals and works to reduce conflict, discord and strife; it also allows each person to make sensible decisions on whether to save, consume, use, or dispose of the property that is committed into his or her hands. That last alternative (which includes the power to sell, lease, bequeath, mortgage, and share) calls for a legal system of voluntary exchange, which lets individuals freely choose their trading partners with whom they can then deal on mutually convenient terms. That same regime of contract also allows individuals to enter into labor and employment contracts. The precise rules of contract enforceability (must there be a writing or other formality?) may differ between legal systems, or between different types of contracts within the same legal system. But some simple exchange mechanism is found in every legal system because it is everywhere needed.

That system of individual liberty, private property, and voluntary exchange places sharp restraints on the use of force and fraud. To natural lawyers, force and fraud negate the independence of the will necessary for individual self-realization. The consequentialist reaches the same conclusion by observing that only voluntary exchanges will consistently produce mutual gain: coerced transactions usually allow one side to profit at the expense of another — hardly a way to achieve social prosperity. To be sure, exchanges will never be truly voluntary if one party uses force or fraud to induce someone to enter into a contract, or takes advantage of the youth, dependence or incompetence of a trading party. So every legal regime of freedom of contract must also develop rules to weed out contracts whose formation is tainted by illicit means.

This basic system of liberty and property helps explain many of the common features of everyday life: why people marry, start businesses, join firms, or sell real estate. It also explains the common prohibitions against trespass, fraud, defamation,

imprisonment and the like. But notwithstanding its impressive achievements, the classical legal theory may be attacked in one of two ways. One is by friendly amendment; the second by frontal assault.

The Friendly Amendment

The friendly amendment begins by observing that force and fraud are not the only obstacles that stand in the path of useful social interactions. To be sure, we all have stirring visions of individuals who hold firm in their beliefs against the multitude and are proven right with time. But in some cases, an individual decision to hold out against the common good produces undesirable results for everyone within the group. First note that the rights of liberty and property are not self-enforcing; someone must supply the state with the wherewithal to raise funds to secure these rights for individuals. But no system of public finance can rely on voluntary donations, for each (self-interested) person would happily free ride on the efforts of others without contributing anything to the central system of enforcement. When such a practice becomes universal, the system of governance fails and individuals are thrown back into hostile and unstable social environments that all would like to avoid.

To overcome these coordination problems (or as they are sometimes called, holdout or public goods problems), the law must develop a sensible system of coercion. More concretely, the classical liberal model must resort to taxation, whose scope should be limited to prevent self-interested public officials from converting tax revenues into forced redistribution. Most classical writers followed Locke in their preference for proportionate taxation as the best curb against official abuse. Deciding who should be taxed, by what formulas, and for what ends set out some of the central challenges to the traditional regimes that honor liberty and property.

The challenges go beyond this as well. Any well-run society needs a system of public infrastructure to support the various activities of a market economy. But once again, it is difficult to see how purely voluntary contracts can provide for public streets, sewers and power. Private property works well when concentrated plots are used for factories, homes and farms. It is more difficult, however, for private property to provide the transportation and communication networks that link these separate holdings together. Sometimes the common property needed for these connections is forged by nature: rivers and oceans are natural modes of communication that are open to all, and within the classical framework, no individual can convert these resources to private use by the rules of occupancy and capture that apply to land, animals and chattels. But in other cases, the networks must be created by state condemnation coupled with the payment of just compensation to the former owner. Sometimes that compensation is paid in cash raised from general revenues, but other times it is supplied in kind: the access to a public highway could easily increase the value of a person's retained lands so that the two together are worth far more than the original, isolated parcel.

The creation of these vast networks (such as common carriers, public utilities, etc.) also raise questions as to their operation. One possibility is that the owner of a network, either public or private, enjoys the usual incidents of ownership, including the absolute right to exclude. But however desirable that power might be in a

competitive industry, it is far more dangerous when a single person controls access to public roads, railroad service or electrical power. In response to this show of monopoly power, the standard liberal theories recognized a separate class of property, "affected with the public interest," over which the state could impose some nondiscriminatory and reasonable system of rate regulation. The right to sell to whomever one wanted at whatever price one wanted never took hold with these key elements of social infrastructure, even when owned by private citizens.

This concern with monopoly also expressed itself in the legal willingness to distinguish contracts in restraint of trade between rival sellers from ordinary contracts of sale between producers and consumers. At the very least, most legal systems refused to enforce contracts to fix prices or to divide markets by product line or by territory. And in many cases, aggrieved customers are also provided with a direct legal remedy for damages or an injunction. The exact scope of the antitrust law (or, as it is called outside the United States, competition policy) has been subject to extensive dispute. But there are relatively few political thinkers willing to profess complete indifference to the problem of market power. Some greater form of state intervention is at least tolerated, and perhaps welcomed, even within the classical legal framework.

State power is also extended in a second dimension. Sometimes the protection of life, liberty and property must be done after the fact. While money damages are available to individuals whose rights are violated by another, such compensation is an imperfect remedy for death, or serious bodily harm, or even the loss of one's home and possessions. The state police power may therefore be invoked to prevent these events by imposing restrictions on what people may do. Individuals may be required to get a driver's license to use public roads; or a state permit to drill in crowded areas; or be banned from keeping wild animals in private homes. The current disputes over gun regulation offer the most vivid contemporary illustration of the problem. The case would be easy if only criminals used guns. But it becomes far more complex when most — perhaps 99 percent or more — guns are used for legitimate purposes, including the prevention of violence. Generally speaking, in each of these cases, a state restriction may cut out the legitimate exercise of a personal liberty, but the failure to impose that restriction may result in the occurrence of an avoidable but noncompensable harm. The ostensible certitude of the system of rights is thus undermined by the uncertainty of what the future holds in store. All legal systems must therefore decide what conduct should be banned, what should be regulated, and what should be left untouched. And every system is forced to minimize the overall errors that come from both too little, and too much regulation.

The Frontal Assault

The friendly amendments to a system of liberty and property thus allow public taxation, public infrastructure, and public regulation of dangerous private activities. Even if the state undertakes all these tasks, the preservation of liberty and property remain its central goal. The frontal assault on the classical legal system takes a very different view. It treats the amendments to the system of liberty and property as conclusive evidence that these principles do not articulate some self-contained social objectives to which

everyone adheres. From there, the critics go on to say that the gap between protecting liberty and property and advancing a coherent social system is far greater than the classical defenders acknowledge, even after the friendly amendments are taken into account.

The attack is directed toward each and every element of the classical synthesis. Autonomy is no longer regarded as the necessary bulwark of individual identity. Individual talents are not considered to be earned in any strong sense of that word; rather, they result from a mix of good genetic fortune coupled with the favorable circumstances of birth. To privilege the classical liberal system's distribution of rewards and punishments simply compounds the vagaries of nature's lottery and social life rather than counteracting them. Accordingly, taxation is no longer limited to financing public goods that make the utilization of private property more efficient. It now has the additional function of redressing any imbalance of fortune and wealth brought about by the circumstances of birth and the unjust set of social institutions that privilege a select few.

By like reasoning, the origins of property are no more sacrosanct than the origins of liberty. To the critic, taking possession of an unowned thing is not some uncontroversial unilateral act that establishes a relationship between a person and a thing. Why the accidents of speed, location and luck should be dispositive is no clearer for the distribution of natural resources that the world has to offer than it is for human talents. Some system of collective allocation could therefore regulate the acquisition of property, even when concerns with public goods and monopoly are nowhere to be found.

Voluntary exchange also comes under prolonged scrutiny. The exceptions for coercion and duress are, under the classical vision, limited to the use of force and fraud. They do not, absent monopoly, include cases where the inequality of economic wealth and social power skew a transaction in favor of those who already occupy a position of privilege in society at large. The persistent critic of classical liberalism is quite willing, if not eager, to impose substantive limits on the kinds of contracts that ordinary people can enter: usury laws, minimum wage laws, and safety laws are one set of responses to the perceived social imbalance. In other cases, powerful parties may lose their ability to choose the individuals with whom they deal: thus antidiscrimination laws prevent employers (who have economic might) from discriminating against prospective or current employees on grounds of race, sex, national origin, age, handicap and, in some jurisdictions, sexual orientation.

The scope of the tort law is not beyond challenge either. To be sure, few people argue for relaxing the traditional prohibitions on the use of force or fraud. But the difficult question is whether to expand the recognized harms beyond these. In some instances, the law might allow protection against economic harms caused by "ruinous competition" or market predation; or it could require powerful firms to deal with any willing provider of certain goods and services. These concerns lead to the passage of protective tariffs, antidumping laws in international trade, the manifold restrictions on entry that have been found in the banking, insurance, railroad and transportation industries in our own time, and the extensive regulation of the types of contracts that health care providers can provide to their patients.

The same determined critique spills over to the friendly amendments to the liberty and property regime. So long as the current distribution of natural talents and private property are suspect, why use proportionate taxation, which does nothing to redress the imbalances created by the classical legal rules and the market transaction they protect? Progressive taxation on income, estate, gift, and inheritance is now justified as a powerful instrument for the redistribution of wealth from affluent to needy individuals. Likewise, public regulation changes its focus from insuring the smooth operation of networks and the prevention of the use of monopoly power to building explicit redistributive notions into the system. Examples include guaranteed access to health care at subsidized rates for the poorest members of society, or, explicit subsidies for schools and hospitals and other preferred social institutions, whether through subsidized bus rates or free Internet access.

Finally, the scope of regulation under the police power expands. No longer is the system of regulation tied to preventing various kinds of common law torts. Now, in addition to preventing pollution and disturbance, state land use regulation can impose height limitations, street setbacks, density limitations, growth moratoria and the preservation of natural habitats for endangered species, all without having to compensate the owner for his loss of property value attendant on those actions, unless the entire value of the land is destroyed. The scope of public regulation increases as the rights of liberty and property decline. In their place emerges a system of positive or welfare rights where all individuals have some minimum claim on the state (or their fellow citizens) for some fraction of the social wealth they produce.

No final resting place

Naturally enough, the new program of increased government regulation and redistribution is not beyond challenge by the defenders of liberty and property. They have been buoyed by the failures of socialism and communism around the world, and the economic miracles in Third World countries that follow market principles. To be sure, in some ideal world, a system of "from each according to his ability, to each according to his need," might guide the distribution of goods and services in society. But before these goods can be fairly distributed, they must be produced. The rules that separate reward from production undercut the incentives of self-interested individuals to produce needed goods and services in the first place. And the willingness of the state to go beyond the traditional norms of taxation and regulation hardly explains why a progressive tax makes sense, or how much land use regulation should be imposed and why. Each form of regulation invites private conduct that will negate its intended consequence. High taxation leads to underreporting or even emigration. Strong habitat controls can lead to the premature destruction of valuable trees and endangered wildlife by landowners seeking to preempt the state's regulatory power. But with each possible evasion comes some determined social response to control the loophole: stronger penalties against tax evasion; prohibitions on emigration; government inspections and criminal punishment to prevent habitat degradation.

In the face of this controversy, political sentiment tends to move to some undetermined middle point. But the range of intermediate accommodations that respect

in varying proportions both market power and state authority is very broad. Even in the past 20 years, the mixed solutions of a Ronald Reagan differ markedly from those of a Bill Clinton. And so the cycle continues. Where a well-run society should rest on the continuum between public and private control has been the most contentious and thorny issue of legal and social theory throughout the generations. Hopefully, some answers to this ongoing dispute are contained in the five volumes of material assembled here.

Volume Introduction

The standard classification of common law topics recognizes the sharp division between the rules of property and those of contract. The former are concerned with the acquisition and definition of property rights in both tangible and intangible things. So stated, these rules perform the essential task of matching particular resources with particular individuals, and thus prevent social life from degenerating into a perpetual free-for-all that utterly undermines the stability of possession. Yet it would be a mistake to assume that the sole function of legal rules is defensive. The individual who has secured the safety of his person and possessions may be ill-equipped to use them efficiently; or he may wish to make gifts of what he owns to family or friends. To facilitate these objectives some rules for voluntary transfer are required. The rules concerning gifts generate little controversy in to the operation of the overall legal system, for there is scant reason to think that donors have systematically taken advantage of donees. But most transfers of capital and labor, especially those between strangers, are not done from some disinterested sense of benevolence, but in the expectation of receiving something greater in return. It becomes, therefore, critical to assess the uses and limitations of the system of voluntary exchange that allows for the transfer of both services and property. The essays in this volume address both the uses and limitations of contractual exchange. In so doing, they explore the larger issue of freedom of contract, which has long been a lightning rod for both proponents and opponents of laissez-faire. At the same time, they explore some of the internal complexities of the law of contract, which, when understood, aid in understanding the grander legal doctrines.

The first essay in this volume, Morris Cohen's *The Basis of Contract*, addresses both the large and small questions of contract law. Cohen reminds us forcefully that contracts involve not only private volition but state coercion, and he takes us on a quick tour of the use of contract in both ancient and modern societies. As befits the topic, he then asks which promises should be enforced. In one sense, only bargains that involve a transfer of equivalents between two parties should be enforced, a point which leads him to question whether modern labor contracts should be respected given what he perceives as the inability of the worker to turn down contract provisions. Yet later in his article, Cohen returns to more classical themes of contract law as it relates to the doctrines of consideration and form. The court that seeks to examine the equivalence of what is exchanged may become so intrusive that it blocks the development of the

very commerce that it wishes to advance. Yet by ignoring the relative values of the two parties, it reduces the requirement of consideration to a mere "form" that could open the way for all sorts of advantage-taking that have no sensible place in a market economy. Finding a path through that dilemma raises problems to this very day.

Friedrich Kessler's highly influential essay, *Contracts of Adhesion*, moves from technical contract doctrine to matters of high politics. Kessler, a gifted technical lawyer, starts with concrete illustrations of defects in the formation of insurance contracts: what should be done if a loss occurs after the insurance company promised coverage, but before it issues a formal written contract? Kessler notes the ingenious ways that courts manipulate legal doctrine in order to supply the missing coverage. Kessler notes that the courts that adhere to freedom of contract principles in form often sidestep them in practice by allowing a tort remedy on the theory that the insurance companies owe the public at large a general duty to act promptly in processing applications. But in examining this situation, he does not address the counterargument that it is in the interest of the insurance carrier to start coverage at the earliest possible date, so long as it can collect the needed premium, and this it can do by issuing immediate binders when coverage is agreed orally. Even today, courts in insurance disputes regularly invoke the idea that the terms of a disputed or ambiguous contract should always be construed in ways that honor the reasonable expectations of the insured who signed the agreement, instead of the insurer who drafted it. Kessler then draws larger implications from his insurance example. Writing during World War II, he presses the point that the power of large private enterprises often matches that of fascist governments because their economic might allows them, as it were, to legislate by contract free of democratic supervision. Yet it is at just this point that the defenders of laissez-faire howl in protest. To be sure, sometimes courts are need to smooth over the rough spots in contracting, but so long as individual consumers have a choice of suppliers, the defenders of laissez-faire contend that it is a mistake to paper over the difference between market choice and state power with the elusive term "adhesion." One can say "no" to General Motors, but not to being drafted by the United States Army.

Kessler's provocative ideas are capable of expansion. John Dawson invokes the idea of economic duress to increase the scope of government intervention into private contracts. In ways that echo Hale's famous essay, Dawson insists that duress covers more than a gun to the head. One instance of the practice is the refusal of a tailor to return a garment unless its owner agrees to pay a sum in excess of the agreed upon price. In dealing with these cases of "duress of goods," the presumptive wrong comes from putting the customer to the choice between the return of his garment or the retention of his money, when he is entitled to both. The situation is complicated because sometimes the tailor might be able to justify his action by pointing to some unanticipated and unavoidable cost that he has been required to bear. This renegotiation problem, as it is now called, is one of the central problems of a system of contract. The key question, however, is whether Dawson is correct insofar as he thinks that the inequality of bargaining power between workers and employers, for example, is yet another form of duress that justifies state intervention. Once again the reply is that so long as workers have a choice of employers, then the coercion metaphor is inapposite. And in the few cases where they do not, the real question is not coercion proper, but the use of

monopoly power without some offsetting obligation of good faith.

The question of monopoly power is the central theme in William Letwin's historical essay on the common law concerning monopoly, or contracts in restraint of trade. The issue here does not, at least by traditional argument, raise questions of coercion or duress, but it does raise questions of the proper allocation of resources. To the die-hard libertarian, one contract is as good as another so long as neither involves the use of force or fraud. An agreement to sell a bushel of grain is indistinguishable, on this point of view, from a contract to cartelize the grain market; after all the content of a bargain is not the concern of the state. The common law never took that uncompromising position, but instead held that cartelization contracts were unenforceable in virtue of the fact that they tended to raise prices and reduce output in ways that retarded social welfare. The economics of monopoly still lay in the future, but as these issues became clearer toward the end of the nineteenth century, the rise of the antitrust movement in the United States rested in large measure on the perception that the monopoly threat could not be countered by the single (if cheap) legal response of refusing to enforce these contracts. In addition, the law provided for criminal sanctions against these contracts, and treble damage actions from the victims of monopoly power. The scope and reach of the antitrust laws remains a central concern today for both the critics and defenders of laissez-faire.

The concerns over the enforcement of contract also expressed itself in the doctrine of unconscionability, which was explored in Arthur Leff's article *Unconscionability and the Code*. The Code itself is restricted to contracts of sale, including sales to consumers who are ignorant of the basis of market transactions. But the issues in question go far beyond, to all sorts of contracts. As befits the subject, the rules of "procedural" unconscionability address the question of whether a skilled party (say a ghetto merchant) has gained a contractual advantage by unfair surprise, when he includes obscure terms within the contract or diverts his customer's attention from any onerous terms. Broad rules of unconscionability allow courts to set aside these contracts without having to prove some specific element of fraud that seems likely under the circumstances. Substantive unconscionability refers to terms that seem so lopsided that no rational person could have entered into them. The doctrine of unconscionability thus goes counter to the general common law rule that the content of a bargain, at least in ordinary consumer transactions, is of no concern to the courts. The hard question is to decide whether the use of this doctrine causes more confusion than it eliminates, by casting doubt on transactions that do make commercial sense, at least to a court sophisticated enough to understand them — a matter on which Leff joyously speculates. In more recent times the unconscionability doctrine has been in judicial retreat, largely because of the easy access to large chain stores, catalog sales, credit cards, and more recently, e-commerce. Yet at the same time many specific statutes, such as those requiring "cooling off" periods, have been introduced to respond to real or perceived abuses. Yet once again the question remains: do these statutes create more problems than they solve?

The question of judicial intervention is not confined to contracts of sale. At common law, the standard default rule for employment contracts was that the employer could fire a worker for good reason, bad reason, or no reason at all. Contracts for

"permanent" employment were really treated as contracts at will, so that the employee who desired protection against unjust dismissal had to persuade the employer to offer a contract for a term of years, or specific protections against wrongful dismissals. Lawrence Blades led the attack against the contract at will by positing that these arrangements worked against the individual freedom of workers given the risk of employer power. The danger of arbitrary dismissal was such that Blades became an early champion of the modern common law rules that give workers protection against wrongful dismissal, often by allowing huge damages for emotional distress and economic loss. The obvious disparity in wealth between the parties was said to justify the judicial intervention in a manner that Cohen, Kessler, and Dawson would approve.

I have taken the opposite position in my subsequent defense of the contract at will. Once again the nub of the issue is: what matters, the relative net worth of the employer or the employee, or the number of alternative opportunities available to each? Surely, the latter must make some difference, for no person is indifferent to the prospect of having only a single employer in a given town or trade. The presence of alternatives forces an employer to meet the terms and conditions offered by others, which is why wages in high-skilled industries are higher than those in low-skilled industries. The possibility of quitting one job and finding another (which occurs more often than dismissal, with or without cause) has strong efficiency properties. Any employer who tries to take advantage of his workers reduces the net gain the workers obtain from the job, and thus the employer runs the increased risk of worker defection. Given the costs of training new workers, and the risk to reputation, powerful economic incentives keep employers in line wholly without any expensive judicial term that treats all contracts as terminable only for cause. Today common law courts are split on the merits of the contract at will; those states that abrogate the rule often set very different substantive rules for defining the common law wrong. Many legislatures have gotten into the act by imposing restrictions on the power to hire and fire at will. In the 1930s, the collective bargaining statutes were an explicit repudiation of the common law doctrine. Today the antidiscrimination laws, based on race, sex, age, and handicap, have, if anything, far passed the rules of collective bargaining in importance. Deciding whether to embrace or reject these statutory fixtures turns in large measure on the attitudes that courts, legislatures and citizens take to the contract at will.

The concerns with freedom of contract are not only expressed in terms of market exploitation. Margaret Radin's exploration of the issue of "market-inalienability" asks whether the state should impose restrictions on the ability of individuals to enter into certain kinds of transactions for cash. Is it permissible for one to sell one's body, one's body parts, or one's baby? In Radin's view, there are social concerns with commodification, or treating special relationships as fungible commodities when they are not. It may be appropriate for one to donate kidneys or blood, but it is quite a different proposition to allow their sale, given what it tells us about human beings. In raising these objections, Radin takes explicit issue with the narrower justifications for restrictions on market alienation that are more congenial to the defenders of laissez-faire. Sales of organs, for example, raise the risk of abuse; sales of babies raise obvious questions about the welfare of individuals who do not have the capacity to contract on their own behalf. The key question is which approach better explains what the law

is and what the law ought to be. Do we think that it is inappropriate to allow surrogate motherhood, or do we have confidence that the surrogate mother can be separate from her child in ways that allow often desperate couples to have a child of which the husband is the father?

Lisa Bernstein's article has a different take on the law of contract. Instead of looking at contracts as isolated transactions devoid of institutional structure, she examines the operation of exchange arrangements in one specific market: the market in diamonds dominated (at least before her article was written) by Chasidic Jews in New York. Bernstein notes little private demand for judicial enforcement of contracts. Instead, the parties who are engaged in repeat dealing prefer to rely upon local arbitration before industry experts who pay little attention to the formal rules of the common law and much to the specific guidelines that they have drawn up to govern themselves. The efficiency of these rules stems from the close personal interaction between the parties, their common religious background, and the reputational sanctions available against those who deviate systematically or deliberately from the local norms. Bernstein's study gives some hope to those who think that organized markets have ways to avoid the injustices that are sometimes found in isolated judicial transactions. The question that we have to ask is whether, and to what extent, these norms and practices can survive when the diamond market extends from Saudi Arabia to Japan. Will the older informal institutions adopt to take into account the new environment, or will the parties be forced to rely more on imperfect mechanisms of judicial enforcement? The questions here raise issues central to the future of private contracting, and to the soundness of the positions of both those who welcome, and those who fear, judicial enforcement of (not to say judicial interference with) private contractual arrangements.

This page intentionally left blank

HARVARD
LAW REVIEW

VOL. XLVI FEBRUARY, 1933 NO. 4

THE BASIS OF CONTRACT *

THE nature of contract has been much discussed by lawyers interested in specific technical doctrines, and by moralists, economists, and political theorists interested in general social philosophy. There is still need for some effort to combine these points of view. The bearings of general philosophy become more definite through its applications, and the meaning of a technical doctrine receives illumination when we see it in the light of those wider ideas of which it is the logical outcome.

This large and important task is obviously beyond the limits of a short paper. But a few suggestions may indicate something of the scope of the problem.

I. THE SOCIAL ROOTS OF CONTRACT

One of the most influential of modern saws is Maine's famous dictum that the progress of the law has been from status to contract. It has generally been understood as stating not only a historical generalization but also a judgment of sound policy — that a legal system wherein rights and duties are determined by the agreement of the parties is preferable to a system wherein they are determined by "status."

This easy assumption, that whatever happens to be the outcome of history is necessarily for the best and cannot or ought not to be counteracted by any human effort, is typical not only of the histori-

* This article will appear as part of a book on LAW AND THE SOCIAL ORDER, to be published by Harcourt Brace & Co.

cal school of jurisprudence since Savigny, but also of the general progressive or evolutionary philosophy of Maine's generation and largely of our own. Accordingly, pleas that under present conditions we need certain limitations on the freedom of contract have encountered the objection that we must not go against history and thereby revert to barbarism.

A. Contract in History

Before considering the validity of the last argument let us briefly consider Maine's dictum from the point of view of the present state of historical learning. For while the study of the past in itself is not sufficient to determine desirable policies for today, it is necessary to view reigning ideas in their perspective and past careers if we are to separate them from their obsolete elements. In any case, objections based on inadequate history can be fully met only on the basis of sounder knowledge.

That Maine's generalization is not a universal and necessary law, he himself recognized in his treatment of feudal land tenure.[1] The rights and duties of sovereign and subject, of homage or fealty and protecting lordship, were contractual in the early Middle Ages, and gradually ceased to be so as they became customary and were later replaced by the legislation of the modern national states. It is also true, as Dicey, Hedemann, Charmont, Jethro Brown, Duguit, and Pound have shown, that the modern state has, in all civilized countries, been steadily increasing the scope of its functions, so that men now do things by virtue of their status as citizens and taxpayers which formerly they did by voluntary agreement. One only needs to mention the fields of charity and education to make this obvious. Moreover, in many relations in which men are more or less free to enter into contracts, such as that between insurer and insured, landlord and tenant, employer and employee, shipper and carrier, the terms of the agreement

[1] MAINE, ANCIENT LAW (6th ed. 1876) 170, 305. In referring to the feudal centuries as "the golden days of 'free,' if 'formal,' contract," Pollock and Maitland assert that in that period ". . . the law of contract threatened to swallow up all public law. . . . The idea that men can fix their rights and duties by agreement is in its early days an unruly, anarchical idea. If there is to be any law at all, contract must be taught to know its place." 2 HISTORY OF ENGLISH LAW (2d ed. 1898) 233.

are more and more being fixed by law, so that the entering into these relations has something analogous to the entering into the relation of marriage, trusteeship, or public office. The specific rights and duties are not fixed by agreement, though the assumption of the relations is more or less voluntary.

Nevertheless there is enough truth in Maine's observation to warrant a more discriminating attitude to it than that of complete acceptance or complete rejection.

Looking at the matter macroscopically rather than microscopically, there can be little doubt that legally binding agreements or promises play a smaller part in the earlier history of all known peoples. The development of contract is largely an incident of commercial and industrial enterprises that involve a greater anticipation of the future than is necessary in a simpler or more primitive economy. In the latter the solidarity of relatively self-sufficient family groups and the fear of departing from accustomed ways limit individual initiative as well as the scope and importance of what can be achieved by deliberate agreements or bargains. In some respects, however, less developed societies resort more than we do to contracts or compacts and enforce promises that we no longer enforce. Thus they preserve peace not by organized police or standing armies, but by agreements like our present treaties of peace between nations; and promises to the gods, which are now matters of individual conscience, used to be enforced by the community as a whole because it feared the undiscriminating effects of divine wrath.

It has been assumed since the days of Homer — it is involved in Montesquieu's story of the Troglodytes — that savage men make no compacts or agreements and do not attach importance to promises. This is a doubtful generalization, which modern anthropology does not confirm. We may, however, agree that while commerce in its early as in some of its later stages is connected with war and piracy — " Who but a fool would have faith in a tradesman's ware or his word? " — the main effect of widening markets has been in the long run to favor steady industry and reliance on promises as a basis for individual enterprise otherwise impossible. Extensive commerce, involving travel and contact with other peoples, leads to the observation of habits and ideas different from our own, and thus tends to introduce the disinte-

grating force of rational reflection into the hard crust of traditional mores and beliefs. Customary ideas and familiar ways of doing things thus lose their pristine fixity, and a certain amount of free individual thought and conduct is thus developed, despite the opposition to all innovations expressed by the conservative landed interests. In this state individual bargains or agreements necessarily receive increased attention from the law. This is repeatedly illustrated in the history of the ancient Jews, Greeks, and Romans and of medieval Europe.

The growth of the Hebrew law of contract in the Mishnah seems to have followed the expansion of commerce that came with the capture of their first seaport, Jaffa, by Simon Maccabæus. The older Deuteronomic law, that in the Sabbatical year released all debtors, naturally discouraged credit transactions at certain times. This inconvenience was overcome in the time of King Herod by the institution of the *prosbul*, a contract of record by which loans became debts to the court unaffected by the older law.[2] In general Talmudic jurisprudence favored contracts, even in cases where religious scruples might have led to restrictions.[3] It is interesting to note that the notion of individual responsibility, a point in which religion, commerce, and ideas of contract or covenant meet, was first vigorously put forth when the Jews were settled in Babylonia, where they engaged extensively in commerce. I refer to the prophet Ezekiel. The older view held the family, tribe, or nation responsible for the acts of any one individual, whether he was a ruler like Saul or David or an ordinary rapacious soldier like Achan.[4] God visits the sins of the fathers upon the children to the third and the fourth generation. But the experience of transplantation to a foreign land led Ezekiel to the rejection of the older view that if the fathers eat sour grapes, the children's teeth are set on edge. "The righteousness of the righteous shall be upon him, and the wickedness of the wicked shall be upon him."[5] After this came the further reflection that

[2] See Dr. Greenstone's article on *Prosbul*, in THE JEWISH ENCYCLOPEDIA; I SCHÜRER, HISTORY OF THE JEWISH PEOPLE IN THE TIME OF JESUS CHRIST (1898) Div. II, 362.

[3] See the tractate BABA METZIA 94a, in which the prevailing opinion allows some stipulations as to money matters even contrary to the Torah.

[4] 2d Sam. 21, 24, Joshua 7.

[5] Ezek. 18:20.

sin is a voluntary act, an affair of the heart, and not something that can happen to you by involuntary contact with an object that is " unclean " or taboo, or even when with the most worshipful intentions in the world you touch a holy object.[6] The significance for the law of contract of this notion of individual responsibility for voluntary acts is too obvious to need development.

The same expansion of the law of free contracts under a predominantly commercial régime can be seen in the history of Greece. The Hellenic laws of contract seem to have allowed as much freedom of business transactions as any legal system known to us. This is particularly clear in Athens after the change, under Cleisthenes, from tribal organization and after the rapid expansion of commerce that followed in the fifth century B.C. The effect of commerce on the Roman law of contract can be seen in the change from the old rigid rules of the *jus civile* to those of the *jus gentium* and prætor's edict. When, as a result of the Crusades and other influences, European trade began to expand, the law of contract was liberalized by the extensive use of the oath to bind verbal agreements [7] — a procedure to which the Church yielded support only after some reluctance because of Christ's explicit prohibition: " Swear not at all."

The expansion of the régime of contract since the seventeenth century has been intimately related to the modern commercial revolution in northern Europe following the development of trade with India and America. The commercial revolution, even before the industrial one, served to transform a predominantly fixed, land economy into a more fluid and enterprising one on the basis of money and credit. This situation was largely responsible for the decay of the remaining feudal concepts and the reception of the more commercial, Italianized Roman law in the period of the Renaissance. This movement, as Maitland has shown,[8] went so far even in England as to endanger the old common law — a danger from which we were " saved " by the vested interests of the practitioners who were organized in the Inns of Court. Even so, the rapid expansion of the modern law of contract took a di-

[6] 2d Sam. 6:6–7.

[7] *Cf.* ESMEIN, LE SERMENT PROMISSOIRE DANS LE DROIT CANONIQUE (1888) 1 *et seq.*, 37 *et seq.*

[8] MAITLAND, ENGLISH LAW AND THE RENAISSANCE (1901).

rection in England not so widely different from that in Holland and France, and the work of Coke, Holt, and Mansfield is in many respects parallel to that of Grotius, Voet, Domat, Pothier, and the other commentators on modern civil or Roman law.

Maine's observation that the progress of the law is from status to contract is, therefore, partly true in certain periods of expanding trade. But close on the heels of expansion comes consolidation or closer organization; and in the wake of increased freedom of contract we find increased regulation, either through the growth of custom and standardization or through direct legislation. At no times does a community completely abdicate its right to limit and regulate the effect of private agreements, a right that it must exercise to safeguard what it regards as the interest of *all* its members.[9]

These very brief historic observations do not, of course, settle the issue. But they are sufficient to override the supposed veto of history, and to allow us to consider on their specific merits all questions as to the regulation of contract.

The support of Maine's dictum, however, did not come exclusively from legal history. It had its roots in the general individualistic philosophy that manifested itself in modern religion, metaphysics, psychology, ethics, economics, and political theory. Let us begin with the last.

B. The Political Theory of Contractualism

Contractualism in the law, that is, the view that in an ideally desirable system of law all obligation would arise only out of the will of the individual contracting freely, rests not only on the will theory of contract but also on the political doctrine that all restraint is evil and that the government is best which governs least. This in turn is connected with the classical economic optimism that there is a sort of preëstablished harmony between the good of all and the pursuit by each of his own selfish economic gain. These politico-economic views involve the Benthamite hedonistic psychology, that happiness consists of individual states of pleasures and that each individual can best calculate what will please him most. Back of this faith of legal individualism is the modern

[9] 8 HOLDSWORTH, HISTORY OF ENGLISH LAW (1926) 100 *et seq.*

metaphysical assumption that the atomic or individual mind is the supreme reality and the theologic view that sin is an act of individual free-will, without which there can be no responsibility.

The argument that a régime of free contract assures the greatest amount of liberty for all is characteristic of the eighteenth century philosophy of the Enlightenment and is still essential to the faith of Jeffersonian democracy behind our bills of rights. The older Calvinistic argument for government rested on the need of restraining the wickedness of man (due to the corruption of the flesh) by rules and magistrates deriving their power from God. Against this the deistic and bourgeois Enlightenment developed the contrary view, that men are inherently good and that their dark deeds have been due to the corruption and superstition brought about by tyrants and priests. As we get rid of the latter, the original, benevolent nature of man asserts itself and history indeed shows a gradual but steady progress in the direction of freedom. It was natural for the representatives of the growing commercial and industrial interests to view the state, controlled as it had been by landed barons and prelates (lords temporal and spiritual), as exclusively an instrument of oppression, and necessarily evil. But their argument overshot its mark. They forgot that not only industry but also the whole life of civilization depends on the feeling of security that the protection of the government or organized community affords.

The philosophy of freedom or liberty illustrates one of the most pervasive and persistent vices of reasoning on practical affairs, to wit, the setting-up of premises that are too wide for our purpose and indefensible on their own account.

In the fierce fight against the numerous irrational, tyrannical, and oppressive restraints, men jump to the conclusion that the absence of all restraint is a good in itself and indeed the one absolute good in the political field. The error of this cult of freedom is of the same logical type as that of the tradition which it opposes. The latter argues that since our natural impulses are not free from bad consequences, therefore they are absolutely bad and must be made powerless by checks and balances or some other device. Both sets of arguments jump from the perception of what is evil under certain conditions to the affirmation of an untenable absolute. Let us consider the situation from this point of view.

Since what we want generally seems to us good, freedom, as the removal of obstacles to achievement, is a necessary part or condition of this good. But mere freedom as absence of restraint, without positive power to achieve what we deem good, is empty and of no real value. The freedom to make a million dollars is not worth a cent to one who is out of work. Nor is the freedom to starve, or to work for wages less than the minimum of subsistence, one that any rational being can prize — whatever learned courts may say to the contrary.

The tragic fallacy of supposing that mere absence of restraint or of other temporary evil can be an absolute good is poignantly illustrated when men, chafing at oppressive work or company, suppose that mere release will make them forever happy. When this release comes we may find ourselves abjectly miserable, not knowing what to do with ourselves. We then look for some other work that will absorb our attention or other company to fill our interests. So all revolutionists complaining against the oppression of government must as soon as they are successful — indeed, even to attain such success — set up a new government. The new work, the new company, or the new government must prove more congenial or beneficent if we are to escape or mitigate the human inclination to regret the struggle and the pain that brought about the change.

Does the state always and necessarily seek to oppress the individual? Doubtless it is true — and those who talk about democratic government or government by law should note the fact that all government is by individual men, whether they govern as priests in the name of God, as people's commissars in the name of the proletariat, or as judges issuing orders in the name of the law. And when men are in the position of governors they cannot escape seeing the justice on the side of their own special interests when these interests conflict with those of the rest of the community. Thus governments almost always think it necessary to keep and to perpetuate their power. But even so, the interest of the governors is not always contrary to those of the governed. Even a wolf, if we may modify a parable of Santayana's, can, to insure a steady supply of sheep, become a careful shepherd of his flock.[10] The question to what extent the interests of the governors

[10] SANTAYANA, THE LIFE OF REASON (1905) 71.

and the governed coincide and to what extent they differ is an empirical one. It varies under different conditions. In general, it is psychologically false to assert, as anarchistic individualists maintain, that government rests exclusively on the force of the governors. Governors exercise a certain social function, the function of making decisions. This is in fact felt to be too arduous a task by most people; and kingship or authority is thus sometimes thrust on a man as it was on Saul looking for his father's asses. It is thrust on our newspapers by those who wish an authoritative source of decision in matters of taste as to the latest books, plays, dress, proper social deportment.

If the state oppresses some it also serves as a defense to those who can control it. Generally speaking, even the most tyrannical state has some interest in protecting the mass of its people against the ravages of epidemics, against disturbance of the general peace or attacks on life by those inclined to violence. But as more and more people become interested in the government, its action becomes more like that of an assembly of ambassadors where each tries to get the most for the interests he represents and yet there is some intelligence of the fact that the preservation of the common interests is the condition of the special ones being maintained.

In the United States the Jeffersonian democrats fought against the power of the government both in the nation and in their states. They feared that the merchants and the large landowners of the seaboard, who had controlled the older states, would control the central government. To limit the power of government by a system of checks and balances was therefore the way to assure liberty from oppression. So long as the country was sparsely settled and our people remained for the most part a nation of independent freeholders, this was a workable theory. It never, however, was carried out consistently. For the temptation to use the government for positive ends, such as education or the safeguarding of our commercial interests in the West and on the seas, could not be resisted even by Jefferson himself. In the last forty or fifty years the representatives of large industry have invoked this theory of the bills of rights to limit legislative power to regulate industry in the interests not only of the workers but also of the future manhood and womanhood of the nation. The same group, however, that protests against a child labor law, or against any

minimum wage law intended to insure a minimum standard of decent living, is constantly urging the government to protect industry by tariffs. Clearly, the theory of *laissez faire*, of complete non-interference of the government in business, is not really held consistently by those who so frequently invoke it. A government so limited in its powers that it could do no harm would be useless, since it could do no good.

To draw a sharp line, as Mill does, between those acts which affect one person and no one else and those acts which do affect others, is impracticable in modern society. What act of any individual does not affect others?

> Nor knowest thou what argument
> Thy life to thy neighbor's creed has lent

A contract, therefore, between two or more individuals cannot be said to be generally devoid of all public interest. If it be of no interest, why enforce it? For note that in enforcing contracts, the government does not merely allow two individuals to do what they have found pleasant in their eyes. Enforcement, in fact, puts the machinery of the law in the service of one party against the other. When that is worthwhile and how that should be done are important questions of public policy. Since no government is omniscient, some element of discretion in the conduct of affairs must be left to the individuals who make a contract. We cannot rely entirely on regulations made in advance by the legislature, or *post factum* by judges. In fact, how can the element of discretion ever be completely taken away from those who have to transact business? But the notion that in enforcing contracts the state is only giving effect to the will of the parties rests upon an utterly untenable theory as to what the enforcement of contracts involves. Part of the confusion on this point is due to the classical theory which views a contract as always and entirely an expression of the will or of " the meeting of the minds " of those who make it.

C. The Economic Argument for Contractualism

When the political argument is closely pressed, it is found to rest on the economic one that a régime in which contracts are freely made and generally enforced gives greater scope to indi-

vidual initiative and thus promotes the greatest wealth of a nation. Three arguments have been used in defense of this view.

The first was based on the eighteenth century optimism that assumed, on religious and metaphysical grounds, a preëstablished harmony between man's selfish pursuit of gain and the common good.

> Thus God and Nature planned the general frame
> And bade self-love and social be the same

But this was soon seen to be contrary to fact when factory and mine owners began to exploit men, women, and children in a way that a nation like Great Britain could not tolerate. Factory legislation thus followed as a refutation of the optimistic dogma.

The second argument, the psychologic one of Bentham, rested on the assumption that as happiness consists in a maximum of pleasure, and that as each man knows best what will please him most, a contract in which two parties freely express what they prefer is the best way of achieving the greatest good of the greatest number. This argument blandly ignores the fact that though men may be legally free to make whatever contract they please, they are not actually or economically free. The mere fact of litigation, of appeal to the courts for enforcement, proves that the parties did not achieve real agreement or that their compact has not been found to serve the interest of one of the parties. Men in fact do not always know what will turn out to their advantage, and some of them have a talent for exploiting the ignorance or the dire need of their neighbors to make the latter agree to almost anything.

The psychologic argument has therefore been succeeded by the biologic doctrine of natural selection and the survival of the fittest. The old Providence that rewarded the virtuous and punished the wicked became the biologic law according to which the most fit survive and those who are not fit perish. On the basis of this romantic confusion between biologic and moral categories, Spencerian liberals have opposed all intervention by the state to aid the weak in the economic struggle. This indifference to the fate of those who are economically crushed is extolled as a virtue in the Nietzschean motto: Let the weak perish that the strong may survive. But the absolute division of men into the strong and the

weak collapses before critical reflection. Who are the weak and who are the strong? Those who prove weak in a state of anarchic competition may become strong in a state of wise regulation and protection.

The clearest and most convincing statement of the case for the classical theory of free competition is that of Mr. Justice Holmes.[11] Let us, he urges, get behind the fact of ownership, and look at the processes of production and consumption of goods. The men who achieve great private fortunes do not consume very much of this social wealth. Their fortunes denote rather power to control the flow of goods. And who is better fitted to command this process of production and distribution than the man who wins it in the competition of the market? The assumption behind this is that the man who succeeds in winning a fortune (not, it should be noted, the man who receives it by inheritance) has succeeded because he has been able to anticipate the largest effective demand for goods and to organize the most economical way of producing them.

One weakness of this argument is that it ignores the frightful waste involved in competition. The community as a whole ultimately pays the cost, in labor and capital goods (including their extensive sales and advertising forces), of all the economic enterprises that are allowed to compete and fail. Moreover, the greatest profits do not always come with the greatest productivity. There are monopoly profits, like the unearned increment of land value, that clearly do not arise from productivity of the owners, and there are monopoly profits that are swelled by reducing the output, so that fishermen, wheat and cotton growers, and other producers are often advised to do this. Neither can free competition prevent the paradoxical situation that our economic crises repeatedly show, namely, an overstocked food market and general destitution from inability to buy. The latter is certainly in part due to the fact that, under unrestrained competition, wages and the return for the labor of the farmer are not sufficient to enable the vast majority of the people to buy enough of what they have produced. Thus, some of the supposedly greater efficiency of private over public business, to the extent that it involves lower real wages, is detrimental to the general welfare. The latter de-

[11] COLLECTED LEGAL PAPERS (1920) 279 *et seq.*, 293 *et seq.*

pends not only on the mass of production, but also on the kind of goods produced, on the conditions under which men work, and on the ways in which the product is distributed.

For these reasons it is rare nowadays to find any advocates of a régime of free competition except among certain lawyers and judges who use it to oppose regulation of the "labor contract" by the state. The general consensus among business men has demanded the organization of our Interstate Commerce Commission and Federal Trade Commission, our state railway and public service commissions, our state insurance and industrial commissions, and other administrative bodies that limit and regulate certain essential business contracts. Also the great captains of industry are everywhere trying to eliminate free competition. And those who talk about "keeping the government out of business" are the last to desire that the government shall not help or protect, by proper rules, the business in which they are involved. The differences that divide men in this respect concern the questions of what interests should be protected and who should control the government.

D. Contractualism and Religion

The laws of any people, as a rule, receive not only their sanction but also a good deal of their direction from their religion. Not only is the tendency to completely separate law from religion very modern and limited to certain Protestant countries, but also it has never been perfectly carried out in practice. There are many facts to support those who, like the late Justice Brewer, insist that our law recognizes ours to be a Christian nation. A closer view shows that English and American legal ideas have been shaped not only by the Christian tradition but more especially by the effects of the Protestant Reformation. The general philosophy called individualism was certainly strengthened by the religious movement which denied the claims of the historically organized hierarchy, which made salvation depend primarily upon individual faith, which identified the Church with the body of the individual communicants, and which led to setting up individual conscience as the final authority. For our present purpose we must limit ourselves to noting how the category of contract or

covenant was broadened by the Reformation, and more especially by Calvinism.

Max Weber has made popular the idea, previously suggested in part by Thorold Rogers and others, that modern commerce and Protestantism are closely connected in their ascetic ethics, in emphasizing the virtues of industry and thrift, or saving for the future.[12] Some features of this view have been severely criticized by the veteran economic historian Brentano.[13] But still we may say that Calvinism was predominantly the faith of the commercial classes, in opposition to the Catholic Church, which, as a great international landowner or as ministering to agricultural populations, upheld the old order of a land economy. This manifested itself characteristically enough in the question of interest for loans of money and in other phases of mercantile contracts.

Leaving aside the conduct of the Popes as temporal rulers of their Italian states, we can say that the general teachings and policy of the Church were undoubtedly directed to the maintenance of social order. Commerce largely in the hands of Jews and Lombards was viewed as a possible disintegrating force. The clearest and most liberal statement of this attitude is to be found in St. Thomas's tract *De Avaritia*:

"Trade is rendered lawful when the merchant seeks a moderate gain for the maintenance of his household, or for the relief of the indigent; and also when the trade is carried on for the public good, in order that the country may be furnished with the necessaries of life, and the gain is looked upon not as the object, but as the wages of his labour." [14]

Starting with the popular notion that all wealth is the product of labor applied to nature, the Church regarded the demand for interest or payment for the mere lending of money as a form of avarice, one of the mortal sins. Melancthon and especially Cal-

[12] 1 M. WEBER, RELIGIONSSOZIOLOGIE (1920) 30–236; THOROLD ROGERS, ECONOMIC INTERPRETATION OF HISTORY (1888) 84 *et seq.*; KARL MARX, DAS KAPITAL (2d ed. 1872) 750; KRITIK DER POLIT. OEKONOMIE (1859) 128. Similar views had been expressed by E. Le Lavelaye, De Tocqueville, and Guizot. See the article on *Christianity and Economics* in PALGRAVE, DICTIONARY OF POLITICAL ECONOMY.

[13] L. BRENTANO, DIE ANFÄNGE DES MODERNEN KAPITALISMUS (1916) 142 *et seq.*

[14] Quoted by 1 ASHLEY, ENGLISH ECONOMIC HISTORY AND THEORY (4th ed. 1906) pt. II, 391; *cf.* ST. THOMAS, SUMMA THEOLOGICA 2a, 2ae, Q. 77, art. 4; SCHREIBER, DIE VOLKSWIRTSCHAFTLICHEN ANSCHAUUNGEN DER SCHOLASTIK (1911) 89, 105, 145, 158, 193, 200, 230.

vin questioned this attitude, and the Protestant theologians, even in the Church of England, soon followed them.[15] In general, the old canon law and the Doctors of the Church emphasized custom rather than the will or agreement of the parties in determining fair price and other incidents of contract. But the newer religious movement has been more subjective, and has left things more to individual conscience, without elaborating any definite rules that should govern such conscience in the very complicated situations of modern economic life. Casuistry, indeed, fell into such bad repute that the very word has become a term of opprobrium. This allowed lawyers serving the interests of their commercial clients to formulate the doctrine of complete freedom of contract.

E. Contractualism and Metaphysical Individualism

Ever since Plato employed a realistic metaphysics to defend his social philosophy against the individualistic relativism of the Sophists, theories that emphasize the rôle of compact, conventions, or agreements have relied upon nominalistic metaphysics, that is, upon an emphasis on the particular as an atom or individual (the two terms are etymologically equivalent) as against the abstract universals that serve as integrating relations. The atomistic analysis of physical nature that triumphed in the seventeenth century suggested the similar analysis of society as constituted by separate individuals who are united only by compact or convention.

The prevalent idea that there can be no individual responsibility without free will, though most often asserted by modern idealistic philosophies, involves this atomistic metaphysics. For if the individual will is only a part of the world, it cannot be absolutely autonomous but must be subject to influences from the

[15] For Melancthon's views, see 1 ASHLEY, *op. cit. supra* note 14, at 457 *et seq.*; SCHMOLLER, ANSICHTEN 120. Calvin's celebrated letter to Oecolampadius is printed in his EPISTOLA ET RESPONSA (Beza ed. 1575) 355, and is discussed in BOHM-BAWERK, CAPITAL AND INTEREST (Smart tr.) 28, and in 1 ASHLEY, *op. cit. supra* note 14, at 458–60. See Jewell, *Exposition upon the Epistle to the Thessalonians* in 2 WORKS (Parker Soc. pub.) 851 *et seq.* *Cf.* THOMAS WILSON'S DISCOURSE ON USURY (Tawney's ed.). For a general view of the question of usury, see BRISSAUD, A HISTORY OF FRENCH PRIVATE LAW (1912) 385–88, 390; 8 HOLDSWORTH, *op. cit. supra* note 9, at 100 *et seq.*

rest of the world. But the sharp distinction between mind and nature, inherent in this nominalism, is hostile to the possibility of any natural science of man, and this difficulty has been a moot point in continental social philosophy.

In England nominalism has, since the days of William of Occam, Locke, Hume, and Mill, been carried into the field of psychology and has tended to resolve the individual mind into a number of psychic states more or less closely connected by the laws of association of ideas. On this basis, English liberalism effected a compromise between the determinism necessary for a science of man and the concept of freedom necessary for a practical social program. This viewpoint emphasizes the nominalistic analysis as between different individuals and stresses psychologic determinism as between different states of the mind.

Though this compromise, in view of its obvious theoretic difficulties, can hardly be called intellectually respectable, there is behind it some sound perception. The human mind or will is neither absolutely determined nor absolutely free. As a part of the world it is genuinely dependent and determined. But as a real and irreducible entity it has, like all other realities, a realm of relative independence. When we speak of its being influenced, we must recognize that there is an *it* to be influenced. In daily life this shows itself in our craving for the recognition of our personality as an actual factor in human relations.

In the realm of contracts this principle of polarity means that the element of will can never be eliminated from the field of law or from any other social field. But the law as part of a larger world must be regulated and determined by the nature of the larger order of which it is a part.

II. Excesses of Contractualism

As the result of the various forces that have thus supported the cult of contractualism there has been developed in all modern European countries (and in those which derive from them) a tendency to include within the categories of contract transactions in which there is no negotiation, bargain, or genuinely voluntary agreement. Let us consider a few typical situations.

A citizen going to work boards a street car and drops a coin

in the conductor's or motorman's box. This, or the buying of a ticket, is treated as a contract, and courts and jurists speak of its "terms" and of the rights and duties under it. No one claims that there is any actual "meeting of minds" of the passenger and the street railway corporation. There is no actual offer or acceptance — certainly no bargaining between the two parties. The rights and duties of both are prescribed by law and are the same no matter what, if anything, goes on in the passenger's mind, or in the corporation's, if it has any mind. Moreover, the liabilities of the railway corporation to the passenger are in many circumstances exactly the same even if he does not buy any ticket. Obviously, therefore, we have here a situation in which the law regulates the relation between different parties and it is pure fiction to speak of it as growing out of any agreement of the wills of the parties.

A more serious confusion of fact and fiction occurs when we speak of the "labor contract." There is, in fact, no real bargaining between the modern large employer (say the United States Steel Corporation) and its individual employees. The workingman has no real power to negotiate or confer with the corporation as to the terms under which he will agree to work. He either decides to work under the conditions and schedule of wages fixed by the employer or else he is out of a job. If he is asked to sign any paper he does so generally without any knowledge of what it contains and without any real freedom to refuse. For we cannot freely change our crafts, and if a man is a weaver or shoe laster, he is dependent on the local carpet or shoe factory for his livelihood, especially so if he has a family, which is not as mobile as money. The greater economic power of the employer exercises a compulsion as real *in fact* as any now recognized by law as duress. The extreme form of such duress, the highwayman's pistol, still leaves us with the freedom to accept the terms offered or else take the consequences. But such choice is surely the very opposite of what men value as freedom.

Clearly, then, the element of consent on the part of the employee may be a minor one in the relation of employment — a relation much more aptly and realistically described by the old law as that between master and servant. Down to the end of the eighteenth century this relation was in fact regulated by the gov-

ernment. Wages used to be regularly fixed by justices of the peace under the authority of parliamentary enactments, and even the beer that the master was to serve to the servant with his bread had its strength regulated by law. Any demand by workmen for higher wages or any accession to such demands on the part of masters was a violation of the law. Yet courts now speak as if the effort on the part of the state to regulate wages were an unheard-of interference with the eternal laws of nature. As a matter of fact, it was only after the Civil War that the United States Supreme Court invented the doctrine that the "right *to* contract" is property and is thus protected against real government regulation by the Fifth and Fourteenth Amendments of our Federal Constitution. But so widespread has this idea become that few have noticed its radical novelty.

The spread of contractualistic notions shows itself in the tendency to speak of marriage as itself a contract. Now there are, usually, solemn promises exchanged when the marriage ceremony is performed and there may be agreements as to dowry and other property rights. But the specific legal relations of husband and wife are by no means determined thereby. These relations are entirely fixed by law and the parties to it cannot vary its terms, just as they cannot vary the terms of their obligations to any children they may bring into the world. If there is no sense in speaking of the rights and duties between parents and minor children as contractual, neither is there in speaking of the relations of husband and wife as contractual. The fact that an act is more or less voluntary does not make its legal consequences contractual.

The extreme of contractualistic thought was reached when European publicists of the eighteenth century and American judges of the nineteenth century spoke of the social compact or contract as the basis of society and of all law.

I do not want to add to the many (in fact, too many) refutations of the social-contract theory. The critics of the theory seem to me to have ignored the large voluntary element in the formation and continuance of government. This is unmistakably manifest if we consider the governance of international affairs or the formation of many of our own states as well as of our national Union. Moreover, the tradition of Hebrew history and Greek philosophy that bases government and law upon covenant or agreement has

had a salutary influence in challenging all law to justify itself at the bar of reason. It is good to ask of any law whether it is such as rational beings would adopt if they wanted to establish a society. Nevertheless, there are obviously insuperable difficulties in trying to derive all legal obligations from contract. Children have certain obligations to their parents that are not contractual. Indeed, we may well ask, " Why should we obey laws that our ancestors and not we agreed to? " We may even go further and ask, " Why *should* we keep agreements that we made some time ago when we were younger and less experienced or wise than we are now? " If there is any rational answer to either of these queries, it must take the form of indicating some social good or necessity that is served by our keeping our promises. But if so, why may not the same social good or necessity be served by making children obey and at times support their parents, or by making those who hold property pay for sewage, education, and other communal necessities, even if they do not agree? The merits of the issue are not really affected by introducing a fictional contract.

These attempts to stretch the category of contract err in failing to recognize a certain necessary social solidarity, especially that of any generation with its ancestors, as not the outcome but the very basis of contract itself. Indeed, the vast majority of those who formerly held that we were bound by the original contract of our ancestors also believed that we could be justly punished because of the original sin of our ancestors. The analogy of the theologic identity strengthened the legal one.

III. The Justification of Contract Law

A. The Sanctity of Promises

Contract law is commonly supposed to enforce promises. Why should promises be enforced?

The simplest answer is that of the intuitionists, namely, that promises are sacred *per se*, that there is something inherently despicable about not keeping a promise, and that a properly organized society should not tolerate this. This may also be said to be the common man's theory. Learned writers ignore this because of their interest in showing the evil consequences of allowing

promises to be broken. But the intuitionists can well object that to judge the goodness of an act by its consequences is an obvious evasion by postponing the issue. For when we inquire which consequences are good and which are bad, we face the same question over again. If the terms " good " and " bad " have any meaning, there must be some ultimate character of action that makes them so, just as there is some ultimate character or nature of objects that makes them blue or beautiful. To say that the blueness or beauty of an object depends upon " the observer " only means that a complete answer involves an additional factor that " the observer " brings into the case. There can be no sense at all in speaking of the quality of an object that the observer beholds if there is no object or there are no qualities to behold.

Now there can be no doubt that common sense does generally find something revolting about the breaking of a promise, and this, if a fact, must be taken into account by the law, though it may be balanced by other factors or considerations. In any case, let us not ignore the fact that judges and jurists, like other mortals, do frequently express this in the feeling that it would be an outrage to let one who has broken his promise escape completely.

It is not a sufficient answer to the foregoing position to show that common sense is not consistent, that in many cases we approve the breaking of promises,[16] that the promises which we think ought not to be broken depend on the relation in which people stand to us, and that all these factors vary from time to time. I do not always have the same appetite or aversion for the same food, but when I do like or dislike a given dish, that is a fact which is not to be read out of existence because something else was true on another occasion. If, then, we find ourselves in a state of society in which men are, as a matter of fact, repelled by the breaking of promises and feel that such practice should be discouraged or minimized, that is a primary fact which the law must not ignore.

But while this intuitionist theory contains an element of truth, it is clearly inadequate. No legal system does or can attempt to enforce all promises. Not even the canon law held all promises to

[16] See PLATO, REPUBLIC I, 331B; CICERO, DE OFFICIIS I, c. 10, III, cc. 24–25. The canon law did not regard all promises, even under oath, as binding. See DECRETALES OF GREGORY IX, lib. II, tit. 26, c. 27.

be sacred. And when we come to draw a distinction between those promises which should be and those which should not be enforced, the intuitionist theory, that all promises should be kept, gives us no light or guiding principle.

Similar to the intuitionist theory is the view of Kantians like Reinach [17] that the duty to keep one's promise is one without which rational society would be impossible. There can be no doubt that from an empirical or historical point of view, the ability to rely on the promises of others adds to the confidence necessary for social intercourse and enterprise. But as an absolute proposition this is untenable. The actual world, which assuredly is among the possible ones, is not one in which all promises are kept, and there are many people — not necessarily diplomats — who prefer a world in which they and others occasionally depart from the truth and go back on some promise. It is indeed very doubtful whether there are many who would prefer to live in an entirely rigid world in which one would be obliged to keep *all* one's promises instead of the present more viable system, in which a vaguely fair proportion is sufficient. Many of us indeed would shudder at the idea of being bound by every promise, no matter how foolish, without any chance of letting increased wisdom undo past foolishness. Certainly, some freedom to change one's mind is necessary for free intercourse between those who lack omniscience.

For this reason we cannot accept Dean Pound's theory [18] that all promises in the course of business should be enforced. He seems to me undoubtedly right in his insistence that promises constitute modern wealth and that their enforcement is thus a necessity of maintaining wealth as a basis of civilization. My bank's promise to pay the checks drawn to my account not only constitutes my wealth but puts it into a more manageable form than that of my personal possession of certain goods or even gold. Still, business men as a whole do not wish the law to enforce every promise. Many business transactions, such as those on a stock or produce exchange, could not be carried on unless we could rely

[17] REINACH, THE APRIORISCHEN GRUNDLAGEN DES BÜRGERLICHEN RECHTES (1922) §§ 2–4. Kant himself derives the obligation of contract not from promises but from the union of free-will to transfer rights.

[18] POUND, INTRODUCTION TO THE PHILOSOPHY OF LAW (1922) 236, 276.

on a mere verbal agreement or hasty memorandum. But other transactions, like those of real estate, are more complicated and would become too risky if we were bound by every chance promise that escapes us. Negotiations would be checked by such fear. In such cases men do not want to be bound until the final stage, when some formality like the signing of papers gives one the feeling of security, of having taken proper precautions. The issue obviously depends upon such factors as the relative simplicity of a given transaction, the speed with which it must be concluded, and the availability of necessary information.

At various times it has been claimed that mere promises as such received legal force in Hebrew, Greek, early German, and canon law. None of these claims can be justified.

All biblical references to binding promises are either to those involving an oath or promise to God or else they assume, as a matter of course, some formality such as striking of hands and pledge or security.[19] Greek covenants, or agreements, had to be in writing or to be recorded and were not free from other formalities.[20] The binding character of promises could not have been absolute to a people to whom Odysseus was a hero.

Though the great authority of Gierke, following Tacitus and Grotius, can be cited for the view that the early Germans attached great importance to keeping one's word,[21] the evidence collected by such men as Brunner, Von Amira, Heusler, and Brissaud shows that the Germans, like other peoples, held promises binding only if some real object passed hands or some formal ceremony took place.[22] Otherwise, pledge or security was required.

---

[19] Proverbs 6:1–5, 17:18, 22:26; Psalms 15:4; Job 17:3; see *Covenant*, in ENCYCLOPEDIA BIBLICA; 1 NOWACK, LEHRBUCH DER HEBRAISCHEN ARCHAEOLOGIE (1894) 341 *et seq.*, 354 *et seq.*

[20] See PLATO, REPUBLIC I, 555; PLATO, LAWS V, 729E, 874B, XI, 917–18, 920–21; PLATO, CRITO 51; ARISTOTLE, POLITICS 1, 9, 1–11; ARISTOTLE, RHETORIC 1, 15, § 22; ARISTOTLE, ETHICS N. V 2, 1131a2; THEOPHRASTOS in STOBAEUS, FLORILEGIUM 44, c. 22; *cf.* 2 VINOGRADOFF, HISTORICAL JURISPRUDENCE (1922) c. 9; MITTEIS, REICHSRECHT UND VOLKSRECHT (1891) 479 *et seq.*; BEAUCHET, HISTOIRE DU DROIT PRIVÉ DE LA REPUBLIQUE ATHÉNIENNE (1897) 19 *et seq.*

[21] TACITUS, GERMANIA c. 24; GROTIUS, THE JURISPRUDENCE OF HOLLAND (Lee ed. 1926) bk. III, § 52; GIERKE, DEUTSCHE PRIVATRECHT I, 284, III, 2, n.2, 283; GIERKE, SCHULD AND HAFTUNG (1910) 168.

[22] Brunner, *Der Schuldvertrag Bedarfte einer Bestimmten Horbaren und Sichtbaren Form*, in 1 HOLTZENDORFF-KOHLER, ENCYCLOPEDIA (7th ed. 1915) 137; BRISSAUD, *op. cit. supra* note 15, § 362, at 451; 2 POLLOCK AND MAITLAND, *op. cit.*

More substantial is the case for the canon law, which undoubtedly went further than any other system to enforce bare promises. The Council of Carthage in 348 B.C. made all written agreements binding and this later led to the action *ex nudo pacto* before the courts of the Church. But the use of the oath was a distinctive ceremony and as it was binding in conscience, that is, in one's relation to God, it did not always afford relief to the promisee. The latter was at times even compelled by the ecclesiastical judge to release the promisor. And through the extension of the power of temporal rulers, as well as of bishops, to pass on the validity of the promise under oath, the legal effectiveness of the latter was whittled away.[23]

B. The Will Theory of Contract

According to the classical view, the law of contract gives expression to and protects the will of the parties, for the will is something inherently worthy of respect. Hence such authorities as Savigny, Windsheid, Pothier, Planiol, Pollock, Salmond, and Langdell hold that the first essential of a contract is the agreement of wills, or the meeting of minds.

The metaphysical difficulties of this view have often been pointed out. Minds or wills are not in themselves existing things that we can look at and recognize. We are restricted in our earthly experience to the observation of the changes or actions of more or less animated bodies in time and space; and disembodied minds or wills are beyond the scope and reach of earthly law. But while this objection has become familiar, it has not been very effective. The force of the old ideas, embodied in the traditional language, has not always been overcome even by those who like Langdell and Salmond profess to recognize the fictional element in the will theory.

Another line of objection can be found in the incompatibility of the classical theory with the consequences that the law attaches to an offer. Suppose that I offer to buy certain goods from *A* at a

supra note 2, at 185; ESMEIN, ÉTUDES SUR LES CONTRATS (1883) 69; 2 HEUSLER, INSTITUTIONEN DES DEUTSCHEN PRIVATRECHT (1885) 225; VON AMIRA, GRUNDRISS DES GERMANISCHEN RECHTS § 70, in 3 PAUL, GRUNDRISS DER GERMANISCHEN PHILOLOGIE (1897).

[23] BRISSAUD, *op. cit. supra* note 15, § 375.

given price, and, following his refusal, give him a week's time to reconsider it. If I change my mind the next day but fail to notify him, a contractual obligation will none the less arise if five days later he notifies me that he has accepted my terms. Here obviously there is never a moment of time when the two parties are actually in agreement or of one mind. Yet no one denies that the resulting rights and duties are identical with those called contractual. It does not help the classical theory to say that I am under a legal duty to notify *A* (the offeree) and that if I fail to perform this duty in the proper way, the law will treat my change of mind as a nullity, *as if* it had never happened. The phrase italicized indicates that we are moving in the realm of fiction (or better, rights and duties imposed by law) and not in the realm of fact. No one denies that the contractual obligation should attach in this case; but there is in point of fact no actual agreement or meeting of minds. The latter, then, is not always necessary for a legal contract.

The logical inconsistency of the classical theory is not cured if we say that the law protects not the will but the expression or declaration of the will.[24] Suppose that in the case mentioned I make a solemn declaration of the revocation of my offer, or write a letter but fail to communicate it. The law, in refusing to give effect to my declared revocation, is not protecting my expressed will, but is enforcing a duty on me in the interest of the general security of business transactions.

A more important objection to the theory that every contract expresses the consensus or agreed wills of the two parties is the fact that most litigation in this field arises precisely because of the advent of conditions that the two parties did not foresee when they

[24] References to the enormous literature on the controversy between those who hold that there must be a real will or intention to be bound as well as a declaration of the will, will be found in 1 WINDSCHEID, PANDEKTEN (8th ed. 1900) § 75; 1 ENNECCERUS, LEHRBUCH DES BÜRGERLICHEN RECHTS (1913) §§ 136, 155; 1 STAUDINGER, KOMMÉNTAR ZUM BÜRGERLICHEN GESETZBUCH UND DEM EINZÜHRUNGSGESETZE (1912) 434. See also MANIGK, WILLENSERKLAERUNG UND WILLENSGESCHAEFT (1907) 27–150; BINDER, WILLE UND WILLENSERKLAERUNG IM TATBESTAND DES RECHTSGESCHAEFTES, reprinted from (1910) 5, 6 ARCHIV FÜR RECHTS- UND WIRTSCHAFTS-PHILOSOPHIE. In France this was taken up by SALEILLES, DE LA DÉCLARATION DE VOLONTÉ (1901), for criticisms of which, see Lerebours-Pigeonnière in L'OEUVRE JURIDIQUE DE RAYMOND SALEILLES (1914) 399 *et seq.*; BONNECASE, SCIENCE DU DROIT ET ROMANTISME (1928) 229 *et seq.*

entered into the transaction. Litigation usually reveals the absence of genuine agreement between the parties *ab initio*. If both parties had foreseen the difficulty, provision would have been made for it in the beginning when the contract was drawn up. When courts thus proceed to interpret the terms of the contract they are generally not merely seeking to discover the actual past meanings (though these may sometimes be investigated), but more generally they decide the " equities," the rights and obligations of the parties, in such circumstances; and these legal relations are determined by the courts and the jural system and not by the agreed will of the contesting parties.

Planiol and others have argued that while certain effects of a contract may not have been foreseen by the parties, nevertheless these are effects following from the original objective and are therefore the will of the two contractors.[25] But to argue that, because the law fixes certain obligations, you did foresee something that in fact you did not see is a confusion which would be too ridiculous to criticize were it not so prevalent in juristic discussions. The confusion between what exists in fact and what ought to be according to our theory occurs also in other fields of liability. An employer is held liable for the negligence of an agent, even where he may have specifically warned the agent against it. For instance, a man instructs his servant to exercise his two horses in his field, since the animals are too spirited to be taken on the street. The servant takes the horses into the street, where they commit some damage. The master is held liable. Now the theory holds that the man who caused the damage is liable. Therefore, the master, being liable, is declared to be the " cause " of the damage. In truth, however, he is the " cause " because he is liable, and not *vice versa*. So in contracts men are liable for things that they did not actually foresee; and to say that they intended or willed these results is a fiction designed to save the will theory.

The obvious limitations of the will theory of contract has caused a reaction that takes the form of positivism or behaviorism: Away with the whole notion of will! — the only realities are specific acts to which the law attaches certain consequences, that is, if you do something by word of mouth, by writing, or by any other act that someone else takes as a promise, then the latter can,

[25] 2 PLANIOL, TRAITÉ ÉLÉMENTAIRE DE DROIT CIVIL (2d ed. 1912) § 944.

under certain conditions, bring an action. In its extreme form, this appears in what Dean Pound calls the state of strict law, which, like everything called primitive, is always with us. A developed system of law, however, must draw some distinction between voluntary and involuntary acts. Mr. Justice Holmes thinks that even a dog discriminates between one who stumbles over him and one who kicks him. The whole of the modern law of contract, it may be argued, thus does and should respond to the need of greater or finer discrimination in regard to the intentional character of acts. The law of error, duress, and fraud in contract would be unintelligible apart from such distinction.

C. The Injurious-Reliance Theory

Though this seems the favorite theory today, it has not as yet been adequately formulated, and many of those who subscribe to it fall back on the will theory when they come to discuss special topics in the law of contract. The essence of the theory, however, is clear enough. Contractual liability arises (or should arise) only where (1) someone makes a promise explicitly in words or implicitly by some act, (2) someone else relies on it, and (3) suffers some loss thereby.

This theory appeals to the general moral feeling that not only ought promises to be kept, but that anyone innocently injured by relying on them is entitled to have his loss " made good " by the one who thus caused it. If, as Schopenhauer has maintained, the sense of wrong is the ultimate human source of the law, then to base the obligation of the promise on the injury of the one who has relied on it, is to appeal to something really fundamental.

This theory also appeals powerfully to modern legal theorists because it seems to be entirely objective and social. It does not ask the court to examine the intention of the promisor. Instead, the court is asked to consider whether what the defendant has said or done is such that reasonable people generally do rely on it under the circumstances. The resulting loss can be directly proved and, to some extent, even measured. In emphasizing the element of injury resulting from the breach, the whole question of contract is integrated in the larger realm of obligations, and this tends to put our issues in the right perspective and to correct

the misleading artificial distinctions between breach of contract and other civil wrongs or torts.

Nevertheless, this theory is not entirely consistent with existing law, nor does it give an altogether satisfactory account of what the law should do.

Contractual obligation is not coextensive with injurious reliance because (1) there are instances of both injury and reliance for which there is no contractual obligation, and (2) there are cases of such obligation where there is no reliance or injury.

(1) Clearly, not all cases of injury resulting from reliance on the word or act of another are actionable, and the theory before us offers no clue as to what distinguishes those which are. There is, first, the whole class of instances of definite financial injury caused by reliance on an explicit promise made in social relations, such as dinner parties and the like. Suppose I say to *A*, " If you agree to meet my friends and talk to them about your travels in Africa, I will hire an appropriate room in a hotel and give a dinner in your honor." *A* agrees but fails to come, or notifies me too late to prevent my financial loss. Here the law gives me no redress. Cases like these are often said to be properly ruled out on the ground that those who make them do not intend to be legally bound. And doubtless people generally know enough law to know that they cannot collect damages in such cases. But this argument is rather circular, since liability does not generally depend on knowledge or ignorance of the law. Men are held liable in many cases where they do not intend to be bound legally. There are doubtless good reasons why there should be no legal liability for " social " promises; but our theory does not account for them.

Even clearer are those cases where someone advertises goods for sale or a position to be filled, and, when I come, tells me that he has changed his mind. The fact that I have suffered actual loss from relying on this public statement does not in this case give me a cause of action. The law does not help everyone who has relied on the word or act of another.

(2) In formal contracts, such as promises under seal, stipulation in court, and the like, it is clearly not necessary for the promisee to prove reliance and injury. Certain formalities are binding *per se*. Consider also an ordinary agreement to sell something. Suppose that the defendant, who refuses to receive the goods, offers to

prove that the vendor did not expect the deal to go through and had told others that he did not care whether it did or not. Would that be a bar to recovery? Actual reliance, it seems, is not always a necessary element in the case. The reliance of the promisee may be as " constructive " or fictional as the intention of the promisor. Nor does the plaintiff have to prove actual damage through the defendant's refusal to live up to his promise and take the goods. To be sure, where the law recognizes no loss, only nominal damages are usually awarded. But the fact that the plaintiff receives judgment is of practical, as well as of theoretic, importance. Clearly, the law favors the carrying-out of promises even in cases where there is no actual reliance or actual loss from nonperformance.

(3) Finally, the recovery that the law allows to the injured promisee is not determined by what he lost in relying on the promise, but rather by what he would have gained if the promise had been kept. There are obviously many cases where the injured party is substantially no worse after the breach than if the contract had never been made. He has thus not been in fact injured. And yet he may recover heavy damages if he would have gained heavily by the performance of the contract. The policy of the law, then, is not merely to redress injuries but also to protect certain kinds of expectation by making men live up to certain promises.

There can be no question about the soundness of the injurious-reliance theory in accounting for a dominant phase of the law of contract, and the foregoing difficulties may thus seem petty. But they do call attention to fundamental obscurities in the very idea of " reliance " as well as in the criteria of " injury ". The injurious-reliance theory, like others, calls attention to a necessary element but does not give an adequate account of the whole of the law of contract. Its merits become clearer when its claims are properly limited.

D. The Equivalent Theory

Popular sentiment generally favors the enforcement of those promises which involve some *quid pro quo*. It is generally considered unfair that after *A* has given something of value or rendered *B* some service, *B* should fail to render anything in return.

Even if what *A* did was by way of gift, *B* owes him gratitude and should express it in some appropriate way. And if, in addition, *B* has promised to pay *A* for the value or services received, the moral sense of the community condemns *B's* failure to do so as even more unfair. The demand for justice behind the law is but an elaboration of such feelings of what is fair and unfair.

The equivalent theory of contract has the advantage of being supported by this popular sentiment. This sentiment also explains the primacy of *real* contracts.

While a legal theory must not ignore common sense, it must also go beyond it. For common sense, while generally sound at its core, is almost always vague and inadequate. Common sentiment, for instance, demands an equivalent. But what things are equivalent? It is easy to answer this in regard to goods or services that have a standard market value. But how shall we measure things that are dissimilar in nature, or in a market where monopolistic or other factors prevent a fair or just price? Modern law therefore professes to abandon the effort of more primitive systems to enforce material fairness within the contract. The parties to the contract must themselves determine what is fair. Thereby, however, the law loses a good deal of support in the moral sense of the community.[20]

Though legal historians like Ames are right in insisting that the common-law doctrine of consideration did not originate in the law's insistence on equivalence in every contract, the latter idea cannot be eliminated altogether. It colors the prevailing language as to consideration, and especially the doctrine that in a bilateral contract each promise is consideration for the other. If a bare promise is of no legal validity, how can it be of any profit to the promisee or of any detriment to the promisor? Clearly, two things that are valueless cannot become of value by being exchanged for each other. The real reason for the sanctioning of

[20] The old doctrine of *laesio enormis*, which used a plausible standard as to when a disparity of value made a contract unfair, has been abandoned because of the growth of the doctrine that we must let the parties completely determine the values involved. This is a great convenience in the process of adjudication, for it rules out all sorts of material inquiries. But it introduces a certain literal rigidity into the law characteristic of that which Dean Pound calls "the stage of the strict law."

certain exchanges of promises is that thereby certain transactions can be legally protected, and when we desire to achieve this result we try to construe the transaction as an exchange of promises. Consideration is in effect a formality, like an oath, the affixing of a seal, or a stipulation in court.

E. Formalism in Contract

The recognition of the formal character of consideration may help us to appreciate the historical myopia of those who speak of seal as "importing" consideration. Promises under seal were binding (because of the formality) long before the doctrine of consideration was ever heard of. The history of forms and ceremonies in the law of contract offers an illuminating chapter in human psychology or anthropology. We are apt to dismiss the early Roman ceremonies of *mancipatio*, *nexum*, and *sponsio*, the Anglo-Saxon *wed* and *borh*, or the Frankish ceremonies of *arramitio*, *wadiatio*, and of the *festuca*, as peculiar to primitive society. But reflection shows that our modern practices of shaking hands to close a bargain, signing papers, and protesting a note are, like the taking of an oath on assuming office, not only designed to make evidence secure, but are in large part also expressions of the fundamental human need for formality and ceremony, to make sharp distinctions where otherwise lines of demarcation would not be so clearly apprehended.

Ceremonies are the channels that the stream of social life creates by its ceaseless flow through the sands of human circumstance. Psychologically, they are habits; socially, they are customary ways of doing things; and ethically, they have what Jellinek has called the normative power of the actual, that is, they control what we do by creating a standard of respectability or a pattern to which we feel bound to conform. The daily obedience to the act of the government, which is the basis of all political and legal institutions, is thus largely a matter of conformity to established ritual or form of behavior. For the most part, we obey the law or the policeman as a matter of course, without deliberation. The customs of other people seem to us strange and we try to explain them as ceremonies symbolic of things that are familiar or seem useful to us. But many of our own customs can appear to an

outsider as equally non-rational rituals that we follow from habit. We may justify them as the sacred vessels through which we obtain the substance of life's goods. But the maintenance of old forms may also be an end in itself to all those to whom change from the familiar is abhorrent.

F. Contract and the Distribution of Risks

Mr. Justice Holmes has suggested that a legal promise may be viewed as a wager: I assure you of a certain event (which may or may not be within my control) and I pay in case of failure.

This view has not found much favor. The first objection that has been urged against it is that when men make a contract, they contemplate its performance rather than its breach. This is hardly fatal. Men can and do sometimes deliberately plan to pay damages in certain contingencies rather than carry out their legal promises. It might even be said that the law sometimes encourages that attitude. Thus, up to the period of the Reform Bill, English law definitely put obstacles in the way of the lessee of land for a term of years who wanted any relief other than damages. On the other hand, Mr. Justice Holmes fails to dispose of the objection that the law does in some cases — in civil law countries more even than in our equity courts — compel specific performance. Moreover, his theory fails to attain its expressed objective, namely, to dispose of the view that a contract is a qualified subjection of one will to another. For the paying of damages does not flow from the promisor's willingness, but is the effect of the law's lending its machinery to the promisee.

Nevertheless, when taken in a wider sense in connection with Mr. Justice Holmes's general philosophy concerning the risk in all human affairs, his theory is illuminating and important.

All human transactions are directed to a future that is never free from elements of uncertainty. Every one of our ventures, therefore, involves the taking of a risk. When I board a train to go home I am betting my life that I will get to my destination. Now a contract or agreement may be viewed as an agreement for the distribution of anticipated gains or losses. If I agree to sell certain goods or services I expect that I shall be paid in good United States money and that with this money I shall be able to acquire certain

other goods. I do not generally take into account the possibility that the purchasing power of the American dollar may be radically reduced when I receive my pay. That contingency is generally not thought of or else deemed too remote, yet certain bondholders do think of it and specify payment in gold of a certain standard. Now the human power to foresee all the consequences of an agreement is limited, even if we suppose that the two parties understand each other's meaning to begin with. Disputes or disagreements are therefore bound to come up; and the law of contract may thus be viewed as an attempt to determine the rights and duties of the two parties under circumstances that were not anticipated exactly in the same way by the two contracting parties, or at any rate were not expressly provided for in an unambiguous way. One can therefore say that the court's adjudication supplements the original contract as a method of distributing gains and losses.

From this point of view, we may look upon the law of contract as a number of rules according to which courts distribute gains and losses according to the equities of such cases; and the pretense that the result follows exclusively from the agreement of the two parties is fictional. Just as the process of interpreting a statute is really a process of subsidiary legislation, so is the interpretation of a contract really a method of supplementing the original agreement by such provisions as are necessary to determine the point at issue.

If we view the law of contract as directed to strengthening the security of transactions by enabling men to rely more fully on promises, we see only one phase of its actual workings. The other phase is the determination of the rights of the contracting parties as to contingencies that they have not foreseen, and for which they have not provided. In this latter respect the law of contract is a way of enforcing some kind of distributive justice within the legal system. And technical doctrines of contract may thus be viewed as a set of rules that will systematize decisions in this field and thus give lawyers and their clients some guidance in the problem of anticipating future decisions . Thus, for instance, if the question arises as to who should suffer a loss caused by the destruction of goods in transit, the technical doctrine of when title passes enables us to deal with the problem more definitely. In any case, the essential problem of the law of contract is the problem of distribution

of risks. The other phase, namely, the assurance that what the parties have actually agreed on will be fulfilled, is a limiting principle.

IV. Contract and Sovereignty

It was said of Aristotle that whenever he set up a theory he began, like an Oriental despot, by killing off all possible rivals. And this seems to be the fashion in the peaceful world of scholarship. I trust that the foregoing discussion will not appear in that light. It has not been my object to refute the various theories discussed but rather to sift the valid from the invalid elements in them. Not one of these theories logically covers the whole field of contracts. But as they are not mutually exclusive, a more adequate account is possible by utilizing the valid elements of all of them.

This task of formulating a comprehensive theory of contract, that shall do justice to its many sources and various phases, is one that I shall not undertake here. But I wish to emphasize certain considerations that supplement the theories discussed so far.

The cardinal error of the traditional individualistic theories of contract is their way of speaking as if the law does nothing but put into effect what the contracting parties originally agreed on. The best that can be said for this is that it may sometimes be true. But even if that were more generally the case, we should still have to attach more importance to the factor of enforcement than the prevailing theories do. The fact that two people agree to do something not prohibited by the public criminal law and carry out their agreement, or fail to do so, does not of itself bring the law of contract into being. A large number of important agreements, even in business, as in social, political, and religious matters, are left to be directly regulated by other agencies, such as the prevailing sense of honor, individual conscience, or the like. It is an error then to speak of the law of contract as if it merely allows people to do things. The absence of criminal prohibition will do that much. The law of contract plays a more positive rôle in social life, and this is seen when the organized force of the state is brought into play to compel the loser of a suit to pay or to do something. Doubtless most people live up to their promises or agreements either through force of custom or because it is in the long run more

advantageous to do so. But there can be no doubt that the possibility of the law's being invoked against us if we fail to do so is an actual factor in the situation. Even if the transactions that come to be litigated are atypical, their judicial determination is still influential in molding the legal custom. For the ruling in a case that departs from the mode supports or opposes some direction of variation and thus fixes the direction of growth of what becomes customary. The fact, then, that in the general run of transactions people do not resort to actual litigation, is certainly in part due to the fact that they know in a general way what will be the outcome of that process. The law of contract, then, through judges, sheriffs, or marshals puts the sovereign power of the state at the disposal of one party to be exercised over the other party. It thus grants a limited sovereignty to the former. In ancient times, indeed, this sovereignty was legally absolute. The creditor acquired dominion over the body of the debtor and could dispose of it as he pleased. But even now, when imprisonment for debt has been, for the most part, abolished, the ability to use the forces of the state to collect damages is still a real sovereign power and the one against whom it can be exercised is in that respect literally a subject.

From this point of view the law of contract may be viewed as a subsidiary branch of public law, as a body of rules according to which the sovereign power of the state will be exercised as between the parties to a more or less voluntary transaction.

The first rules of public law, generally called constitutional law, regulate the conduct of the chief state officials by indicating the scope of their powers. Within this scope legislatures use their discretion or wisdom to enact certain statutes; and judges, by following precedents, elaborate certain rules as to when and how the power of the state shall be exercised. Among these rules we have the laws of partnership, leases, agreements for services, contracts of surety or insurances, and the like. Now, just as the rules of constitutional law are general and leave blanks to be filled in by the legislature, courts, and administrative officials (whose rules and habitual practices are law to those over whom they have authority), so do the rules of contracts allow men to formulate for themselves, within the prescribed limits, certain rights and duties governing certain transactions between them; and when the parties have thus formulated their agreements, the latter become a part of

the law of the land, just as much as do treaties between our nation and others, compacts between states, contracts between a state or division thereof and a private corporation, or the grant of a pension to the widow of a former president. When a state or a municipality makes a contract with a public service corporation for gas or transportation at a given price to the consumer, no one doubts that such an agreement is part of the legal order. But so are private agreements that the law sanctions. Thus, when a trade union makes an agreement with an association of employers, or even with a single employer, the result is law not only for those " represented " at the signing of the papers but for all those who wish to enter the industry at any time that the agreement is in force. This is in general true of all more or less permanently organized partnerships, companies, corporations, or other groups; and enforceable agreements between individuals, no matter on how limited a scale, are similarly part of the law by virtue of the general rules of state action that apply to them.

If, then, the law of contract confers sovereignty on one party over another (by putting the state's forces at the disposal of the former), the question naturally arises: For what purposes and under what circumstances shall that power be conferred? Adherents of the classical theory have recognized that legal enforcement serves to protect and encourage transactions that require credit or reliance on the promises of others. But we also need care that the power of the state be not used for unconscionable purposes, such as helping those who exploit the dire need or weaknesses of their fellows. Usury laws have recognized that he who is under economic necessity is not really free. To put no restrictions on the freedom to contract would logically lead not to a maximum of individual liberty but to contracts of slavery, into which, experience shows, men will " voluntarily " enter under economic pressure — a pressure that is largely conditioned by the laws of property. Regulations, therefore, involving some restrictions on the freedom to contract are as necessary to real liberty as traffic restrictions are necessary to assure real freedom in the general use of our highways.

From this point of view, the movement to standardize the forms of contract — even to the extent of prohibiting variations or the right to " contract out " — is not to be viewed as a reaction to, but

rather as the logical outcome of, a régime of real liberty of contract. It is a utilization of the lessons of experience to strengthen those forms which best serve as channels through which the life of the community can flow most freely.

Consider, for instance, the position of the man who has to ship his goods. Shall we leave him to bargain with the railroad company? That would certainly not add to the security of business. Experience has shown the necessity of the government's standardizing the transaction in regard to rates and other incidents. Similar considerations hold in regard to life and fire insurance.

The notion that standardization is necessarily inimical to real freedom is a fallacy of the same type as the one that habits are necessarily hindrances to the achievements of our desires. There is doubtless the real possibility of developing bad social customs, as we develop bad individual habits. But in the main, customs and habits are necessary ways through which our aims can be realized. By standardizing contracts, the law increases that real security which is the necessary basis of initiative and the assumption of tolerable risks. Naturally, diverse interests are differently affected by this process of legal standardization. The interests of the railroads, for instance, are not always the same as those of the shippers or commuters. But issues as to justice between conflicting interests in such relations require consideration of the specific factors of the situation. For our present purpose it is sufficient to note that the law of contract in thus dealing with public policies cannot be independent of general political theory. By what has been called the method of judicial empiricism courts sometimes pretend that they avoid facing these issues. What really happens, however, is that they apply old, uncriticized, or unavowed assumptions in their interpretation of the " facts." The advantage of empiricism and of the lack of clear ideas as to the policy of the law is that it sometimes saves the community from the deplorable consequences that would follow if courts consistently carried out some of their professed theories. But the history of human tribulation does not support high expectations from the process of " muddling through."

Contracts are standardized not only by statutory enactments such as the New York legislation on life insurance, by orders of commissions such as the Interstate Commerce Commission and the

like, but also by the process of interpretation that courts apply to human transactions and to their formulated agreements. All agreements, if they are to hold for any length of time, must be constantly revised or supplemented. When disputes arise and courts are appealed to, the latter, by the process of interpretation, do this work of supplementing the existing agreements, just as they generally engage in subsidiary legislation when they interpret statutes. When courts follow the same rules of interpretation in diverse cases, they are in effect enforcing uniformities of conduct.

We may thus view the law of contract not only as a branch of public law but also as having a function somewhat parallel to that of the criminal law. Both serve to standardize conduct by penalizing departures from the legal norm. Not only by decrees of specific performance or by awards of damages, but also by treating certain contracts as void or voidable and thus withholding its support from those who do not conform to its prescribed forms, does the law of contract in fact impose penalties. Thus even when certain practices like gambling, illicit sex relations, or agreements in restraint of trade are not criminal offenses, the law regards them with sufficient disfavor to refuse them the protection of its enforcing machinery.

The function of the law of contract in promoting the standardization of transactions is at all times an important one. And the more developed and complicated transactions become, the more there is need for eliminating as much uncertainty as possible by standardization. This is certainly true today. Consider the case of a man who wants to publish a book, to buy an insurance policy or a letter of credit, to ship his goods or to store them in a warehouse, to lease an apartment, to have gas or electricity or telephone service supplied to him, to mortgage his house, or to obtain a surety bond — in all these and in many other relations his freedom to contract is facilitated by standard forms molded by past law and custom. Naturally, standardized contracts, like other laws, serve the interests of some better than those of others; and the question of justice thus raised demands the attention not only of legislatures but also of courts that have to interpret these standard forms and of administrative bodies that have to supervise their enforcement. In a changing social order these standards or forms must grow or become modified; and to make them function more

serviceably it is not sufficient to wait until trouble develops and is brought before the courts for adjudication. The need of intelligent anticipation that can be effected by initiating inquiries cannot be met by our traditional court procedure, and this has compelled the joining of administrative with judicial power in the hands of bodies like the Interstate Commerce Commission and our various state public service commissions.

A realization of the growth of standard forms suggests the introduction of a point of view in the study of contract similar to what has been called the institutional approach in the study of economics. The classical method in economics starts with a theory of free competition and then seeks to qualify that theory by taking note of the hindrances to the free mobility of capital or labor in actual conditions. While this is perfectly just as a scientific procedure, it postpones an adequate account of actual economic conditions. Recently economists have begun at the other end, that is, with the existing organized social habits involved in economic institutions such as our currency, our technical methods of increasing production, the system of distributing and marketing goods, and the like. From this point of view competition is a real and important incident, but its limitations become more clear in this context. A similar change of approach in the study of the law of contract means beginning not with the bargaining between the two parties, but with the legal form or way of doing things, with the established institution within which negotiation is possible.

One of the most suggestive treatments of the nature of contracts occurs in the few pages devoted to it in Hauriou's *Droit Public.*[27] The French master draws a sharp distinction between contracts and institutions. Contracts are voluntary, fixed, and temporary, while institutions are socially hereditary, grow, and last longer. Yet Hauriou also recognizes that contracts, especially collective ones, grow into institutions. The marriage relation shows the passage from one to the other. Attention to what I have elsewhere

[27] HAURIOU, PRINCIPES DE DROIT PUBLIC A L'USAGE DES ÉTUDIANTS EN LICENCE ET EN DOCTORAT ES SCIENCES POLITIQUES (2d ed. 1916) 196–219. I am also indebted for stimulating reflections along this line to 1 DEMOGUE, NOTIONS FUNDAMENTALES DU DROIT PRIVÉ (1911) c. 4, and to his great treatise, TRAITE DES OBLIGATIONS EN GÉNÉRAL (1923), and to Llewellyn, *What Price Contract? — An Essay in Perspective* (1931) 40 YALE L. J. 704.

called the principle of polarity warns us against making irremediable antitheses or antinomies out of necessary distinctions within any living situation. Nowhere is this warning more necessary than against the absolute separation of freedom of contract from government regulation, the former conceived as purely negative and the latter as purely arbitrary. In actual life real freedom to do anything, in art as in politics, depends upon acceptance of the rules of our enterprise. As has been remarked elsewhere, the rules of the sonnet do not hamper real poets but rather help weak ones. Real or positive freedom depends upon opportunities supplied by institutions that involve legal regulation. Our legislative forces may be narrowly partisan and the rules may be poor ones. But this can be remedied not by the abrogation of all rules but by the institution of better ones.

For this reason the notion that government rests on contract — a notion that runs through both our Hebrew and Greek heritage, and largely conforms to our peculiar American experience — contains a partial truth that should not be utterly disregarded because of some poor arguments in its behalf. If we discard the notion that *all* organized society began in a voluntary contract — a proposition that few have advanced as a literal truth — we may yet recognize that as men become more enlightened they can treat government as if it were a contractual affair, that is, judge the services of governmental rules by the price we pay for them. The great men who founded the rationalistic legal and political tradition of the Enlightenment, Althusius, Grotius, Leibniz, and Locke, may have underestimated the force of tradition, but in treating governmental rights and duties under the categories of contract, they helped to liberalize and humanize our international and our criminal law, as well as the law of private and commercial transactions.

There is no inherent reason for rejecting the view that the roots of the law of contract are many rather than one. Agreements and promises are enforced to enable people to rely on them as a rule and thus make the path of enterprise more secure; but in this connection the law must also go beyond the original intention of the parties to settle controversies as to the distribution of gains and losses that the parties did not anticipate in the same way. Some recognition must always be given to the will or intention of those

who made the contract, but the law must always have regard for the *general* effects of classes of transactions, and it cannot free men from the necessity of acting at their peril when they do not know the consequences that the law will attach to their acts — and this needs to be emphasized in any attempt to formulate a rational theory. The law is a going concern and like all social institutions is governed by habit. It therefore will continue to enforce promises and agreements, for no better reason than that they have been enforced and there is no sufficient countervailing consideration to force or justify a break with the established habit that has become the basis of social expectancies. Legal and other habits are not always deliberately formed to serve a definite purpose. Certain forms or ceremonies arise under special circumstances but continue to appeal to us through the principle of economy of effort: it is generally easier to use the existing forms than to break with them and adopt new ones. Of course old forms may become inconvenient or positive hindrances. They are then whittled away by pious fiction or violently changed by revolutionary legislation. In general, however, the ancient truth that men are creatures of habit will put us on guard against the vain assumption that we can get rid of formalism in the law of contract or anywhere else. We may flatter ourselves on getting rid of seal or other ancient binding ceremony. But we must remember that these forms seemed as naturally obligatory to our fathers as the signing of papers or the administering of oaths seem to us today.

In arguing for their indispensability we may recognize that not all forms are perfectly congenial or responsive to the need of the life that pulses through them. And as men become more enlightened they become more ready to discard, as well as employ, diverse instruments or vessels. Wisdom is not attained either by blind acceptance or blind rejection. We need a discriminating evaluation of what exists and what is possible; and this is something to which we can apply Spinoza's dictum: All things excellent are as difficult as they are rare.

Morris R. Cohen.

College of the City of New York.

CONTRACTS OF ADHESION—SOME THOUGHTS ABOUT FREEDOM OF CONTRACT

FRIEDRICH KESSLER

With the development of a free enterprise system based on an unheard of division of labor, capitalistic society needed a highly elastic legal institution to safeguard the exchange of goods and services on the market. Common law lawyers, responding to this social need, transformed "contract" from the clumsy institution that it was in the sixteenth century into a tool of almost unlimited usefulness and pliability. Contract thus became the indispensable instrument of the enterpriser, enabling him to go about his affairs in a rational way. Rational behavior within the context of our culture is only possible if agreements will be respected. It requires that reasonable expectations created by promises receive the protection of the law or else we will suffer the fate of Montesquieu's Troglodytes, who perished because they did not fulfill their promises. This idea permeates our whole law of contracts, the doctrines dealing with their formation, performance, impossibility and damages.

Under a free enterprise system rationality of the law of contracts has still another aspect.[1] To keep pace with the constant widening of the market the legal system has to place at the disposal of the members of the community an ever increasing number of typical business transactions and regulate their consequences. But the law cannot possibly anticipate the content of an infinite number of atypical transactions into which members of the community may need to enter. Society, therefore, has to give the parties freedom of contract; to accommodate the business community the ceremony necessary to vouch for the deliberate nature of a transaction has to be reduced to the absolute minimum. Furthermore, the rules of the common law of contract have to remain *Jus dispositivum*—to use the phrase of the Romans; that is, their application has to depend on the intention of the parties or on their neglect to rule otherwise. (If parties to a contract have failed to regulate its consequences in their own way, they will be supposed to have intended the consequences envisaged by the common law.) Beyond that the law cannot go. It has to delegate legislation to the contracting parties. As far as they are concerned, the law of contract has to be, of their own making.

1. Max Weber, *Rechtssoziologie*, 3 (2) GRUNDRISS DER SOZIALOEKONOMIK (2nd ed. 1925) 413 *et seq.*

Thus freedom of contract does not commend itself for moral reasons only; it is also an eminently practical principle. It is the inevitable counterpart of a free enterprise system.[1a] As a result, our legal lore of contracts reflects a proud spirit of individualism and of *laissez faire*. This is particularly true for the axioms and rules dealing with the formation and interpretation of contracts, the genuineness and reality of consent. Contract—the language of the cases tells us—is a private affair and not a social institution. The judicial system, therefore, provides only for their interpretation, but the courts cannot make contracts for the parties.[2] There is no contract without assent, but once the objective manifestations of assent are present, their author is bound. A person is supposed to know the contract that he makes.[3] "A mere offer imposes no duty of action upon the offeree; there is no obligation to accept or reject or to take any notice of it."[4] If an offeror does not hear from the offeree about the offer, he is free to make inquiries or to withdraw his offer, but he cannot regard silence as an acceptance. Either party is supposed to look out for his own interests and his own protection. Oppressive bargains can be avoided by careful shopping around. Everyone has complete freedom of choice with regard to his partner in contract, and the privity-of-contract principle respects the exclusiveness of this choice.[5] Since a contract is the result of the free bargaining of parties who are brought together by the play of the market and who meet each other on a footing of social and approximate economic equality, there is no danger that freedom of contract will be a threat to the social order as a whole. Influenced by this optimistic creed, courts are

1a. Pound, *Liberty of Contract* (1909) 18 YALE L. J. 454; Williston, *Freedom of Contract* (1921) 6 CORNELL L. Q. 365; Hamilton, *Freedom of Contract*, 3 ENCYC. SOC. SCI. 450.

2. Urian v. Scranton Life Insurance Company, 310 Pa. 144, 165 Atl. 21 (1933); Imperial Fire Insurance Company v. Coos County, 151 U. S. 452 (1894).

3. In the absence of fraud or misrepresentation parties who have put their contract in writing and signed it will not be heard to say that they have not read it or did not know, understand or assent to its contents provided the document is legible however small the print. L'Estrange v. F. Graucob Ltd., 2 K. B. 394 (1934). For American cases see 1 WILLISTON, CONTRACTS (rev. ed., 1936) § 90A.

4. Prosser, *Delay in Acting on an Application for Insurance* (1935) 3 U. OF CHI. L. REV. 39, 45.

5. Coast Fisheries Co. v. Linen Thread Co., 269 Fed. 841 (D. Mass. 1921); Kaufman v. Sydeman, 251 Mass. 210, 146 N. E. 365 (1925). The evolution and gradual restriction of the privity of contract principle with the help of agency (undisclosed principal), third party beneficiary, assignment and tort doctrines expresses our awareness of the growing impersonality of the market and of the social function of contracts. We regard the undisclosed principal doctrine no longer as a wholly anomalous doctrine which ignores the fundamental notion of the common law "that a contract creates strictly personal obligations between the contracting parties." HUFFCUT, AGENCY (2d ed. 1921) 158; Ames, *Undisclosed Principal—His Rights and Liabilities* (1909) 18 YALE L. J. 443. The privity of contract doctrine no longer perfectly insulates the producer from direct liability to the ultimate consumer, as the food and dangerous instrumentality cases illustrate.

extremely hesitant to declare contracts void as against public policy "because if there is one thing which more than another public policy requires it is that men of full age and competent understanding shall have the utmost liberty of contracting, and that their contracts when entered into freely and voluntarily shall be held sacred and shall be enforced by Courts of justice."[6]

The development of large scale enterprise with its mass production and mass distribution made a new type of contract inevitable—the standardized mass contract.[7] A standardized contract, once its contents have been formulated by a business firm, is used in every bargain dealing with the same product or service. The individuality of the parties which so frequently gave color to the old type contract has disappeared. The stereotyped contract of today reflects the impersonality of the market. It has reached its greatest perfection in the different types of contracts used on the various exchanges. Once the usefulness of these contracts was discovered and perfected in the transportation, insurance, and banking business, their use spread into all other fields of large scale enterprise, into international as well as national trade, and into labor relations. It is to be noted that uniformity of terms of contracts typically recurring in a business enterprise is an important factor in the exact calculation of risks. Risks which are difficult to calculate can be excluded altogether. Unforseeable contingencies affecting performance, such as strikes, fire, and transportation difficulties can be taken care of.[8] The standard clauses in insurance policies are the most striking illustrations of successful attempts on the part of business enterprises to select and control risks assumed under a contract. The insurance business probably deserves credit also for having first realized the full importance of the so-called "juridical risk", the danger that a court or jury may be swayed by "irrational factors" to decide against a powerful defendant. Ingenious clauses have been the result.[9] Once their practical utility was proven, they were made use of in other lines of business. It is highly probable that the desire to avoid juridical risks has been a motivating factor in the widespread

6. Sir G. Jessel, M. R., in Printing and Numerical Registering Co. v. Sampson, L. R. 19 Eq. 462, 465 (1875).

7. PAUSNITZ, THE STANDARDIZATION OF COMMERCIAL CONTRACTS IN ENGLISH AND CONTINENTAL LAW (1937) reviewed by Llewellyn (1939) 52 HARV. L. REV. 700; Llewellyn, *What Price Contract—An Essay in Perspective* (1931) 40 YALE L. J. 704; Issacs, *The Standardizing of Contracts* (1917) 27 YALE L. J. 34; RAISER, DAS RECHT DER ALLGEMEINEN GESCHAEFTSBEDINGUNGEN (1936).

8. For a far reaching clause in a sales contract see Hollis Bros. & Co. Ltd. v. White Sea Timber Trust, Ltd., 3 All Eng. R. 895 (1936). Here the seller of timber from a port in the Arctic Circle open for navigation only about twenty-one days stipulated "this contract is subject to sellers making necessary chartering arrangements for the expedition and sold subject to shipments any goods not shipped to be cancelled."

9. PATTERSON, ESSENTIALS OF INSURANCE LAW (1935) *282 et seq.*

use of warranty clauses in the machine industry limiting the common law remedies of the buyer to breach of an implied warranty of quality and particularly excluding his right to claim damages.[10] The same is true for arbitration clauses in international trade. Standardized contracts have thus become an important means of excluding or controlling the "irrational factor" in litigation. In this respect they are a true reflection of the spirit of our time with its hostility to irrational factors in the judicial process, and they belong in the same category as codifications and restatements.

In so far as the reduction of costs of production and distribution thus achieved is reflected in reduced prices, society as a whole ultimately benefits from the use of standard contracts. And there can be no doubt that this has been the case to a considerable extent. The use of standard contracts has, however, another aspect which has become increasingly important. Standard contracts are typically used by enterprises with strong bargaining power. The weaker party, in need of the goods or services, is frequently not in a position to shop around for better terms, either because the author of the standard contract has a monopoly (natural or artificial) or because all competitors use the same clauses. His contractual intention is but a subjection more or less voluntary to terms dictated by the stronger party, terms whose consequences are often understood only in a vague way, if at all. Thus, standardized contracts are frequently contracts of adhesion; they are *à prendre ou à laisser*.[11] Not infrequently the weaker party to a prospective contract even agrees in advance not to retract his offer while the offeree reserves for himself the power to accept or refuse;[12] or he submits to terms or change of terms which will be communicated to him later. To be sure, the latter type of clauses regularly provide for a power to disaffirm,[13] but as a practical matter they are acquiesced in frequently, thus becoming part of the "living law". Lastly, standardized contracts have also been used to control and regulate the distribution of goods from producer all the way down to the ultimate consumer. They have become one of the many devices to build up and strengthen industrial empires.

And yet the tremendous economic importance of contracts of ad-

10. For an effort of the legislature to protect the interests of the buyer of agricultural machinery see NORTH DAKOTA LAWS (1919) c. 238, construed in Palaniuk v. Allis Chalmers Mfg. Co., 57 N. D. 199, 220 N. W. 638 (1928).

11. The word "contract of adhesion" has been introduced into the legal vocabulary by Patterson, *The Delivery of a Life Insurance Policy* (1919) 33 HARV. L. REV. 198, 222.

12. Cole, McIntyre, Norfleet Co. v. Holloway, 141 Tenn. 679, 214 S. W. 817 (1919) discussed by Corbin in (1920) 29 YALE L. J. 441.

13. See the standard form of an application for a life insurance policy, reprinted in PATTERSON, CASES AND OTHER MATERIALS ON THE LAW OF INSURANCE (1932) 819; Robinson v. U. S. Benevolent Society, 132 Mich. 695 (1903).

hesion is hardly reflected in the great texts on contracts or in the Restatement. As a matter of fact, the term "contract of adhesion" or a similar symbol has not even found general recognition in our legal vocabulary. This will not do any harm if we remain fully aware that the use of the word "contract" does not commit us to an indiscriminate extension of the ordinary contract rules to all contracts. But apparently the realization of the deepgoing antinomies in the structure of our system of contracts is too painful an experience to be permitted to rise to the full level of our consciousness. Consequently, courts have made great efforts to protect the weaker contracting party and still keep "the elementary rules" of the law of contracts intact. As a result, our common law of standardized contracts is highly contradictory and confusing, and the potentialities inherent in the common law system for coping with contracts of adhesion have not been fully developed. The law of insurance contracts furnishes excellent illustrations. Handicapped by the axiom that courts can only interpret but cannot make contracts for the parties, courts had to rely heavily on their prerogative of interpretation to protect a policy holder. To be sure many courts have shown a remarkable skill in reaching "just" decisions by construing ambiguous clauses against their author even in cases where there was no ambiguity. Still, this round about method has its disadvantages as the story of the treatment of warranties in life insurance contracts strikingly demonstrates. Courts, when protecting an innocent policy holder against the harshness of the doctrine, did not state clearly that as a matter of public policy an insurance company cannot avoid liability merely because of the falsity of a statement which has been labelled "warranty". They felt that freedom of contract prevented them from saying so. Instead they disguised as "interpretation" their efforts to change warranties into representations.[14] But this makeshift solution tempted insurance companies to try the usefulness of "warranties" again and again.[15]

Society had thus to pay a high price in terms of uncertainty for the luxury of an apparent homogeneity in the law of contracts. Finally, the legislature had to step in. In many jurisdictions warranties have been put on the same footing with representations; in fire insurance, legislation has even prescribed the contents of the standard policy. No such need has arisen with regard to contracts for reinsurance. Here parties of equal skill and bargaining power are dealing with another.

14. *Cf.* Moulor v. American Life Ins. Co., 111 U. S. 335 (1884); Ehrenzweig and Kessler, *Misrepresentations and False Warranty in the Illinois Insurance Code* (1942) 9 U. OF CHI. L. REV. 209, 210 *et seq.*

15. On the shortcomings of the "interpretation" device which results in a constant struggle between draftsman of standardized contracts and courts, see Llewellyn, Book Review (1939) 52 HARV. L. REV. 702, 703.

Although the episode of warranties, because of the intervention of legislation, belongs largely to the past, another well known controversy still lacks a satisfactory solution. Courts have been unable to agree as to who shall bear the risk of "loss without insurance" caused by an unreasonable delay on the part of an insurance company in issuing a policy of insurance for which application has been made.[16] Here again the pious myth that the law of contracts is of one cloth has stood in the path of progress. The courts, because of their reliance on and preoccupation with "interpretation", were lacking experience in handling this situation.

Most courts have felt rather strongly that a recovery in contract is out of the question. According to a "thoroughly established principle of the law of contracts, within the field of which insurance largely lies", an application for insurance is a bare offer and therefore imposes no liability upon the insurance company until it is accepted. Nor does it "afford a basis for any liability by reason of delay in accepting it or the want of care in dealing with it". A decision of the Connecticut Supreme Court has summed up the arguments against an implied contract in the most persuasive form:

> It is of course true that failure to act upon it may, in such a case as this, cause loss to the applicant or to those to be named beneficiaries in the policy, against which he expected to secure protection. That situation is not, however, peculiar to the insurance law; for example, one may make an offer to buy goods which he needs at a certain price, having reason to believe the price will advance, and may incur loss through the failure of the one to whom it is made to act upon the offer within a reasonable time.[17]

To fortify the argument we are told that an implied promise for future action would be unsupported by consideration. "No legal benefit moved from the applicant to it by reason of the offer, and any detriment which the applicant suffers is not one which was contemplated by the terms of the offer or its acceptance". This is all the more true, courts assure us, since the applicant does not agree not to seek insurance elsewhere and is at liberty to withdraw his offer any time before acceptance.[18]

The argument that a recovery in contract would be "contrary to the well settled principles of contract law" has influenced almost the whole of legal literature,[19] particularly since applications typically contain a provision to the effect that the company shall incur no liability under the application until it has been approved by the home office and a formal

16. Prosser, *supra* note 4.
17. Swentusky v. Prudential Ins. Co., 116 Conn. 526, 534 (1933).
18. *Id.* at 534.
19. Only a few decisions have spelled out a contractual or quasi-contractual liability, for instance, Columbian Nat. Life Ins. Co. v. Lemmons, 96 Okla. 228 (1923).

policy issued and delivered.[20] Besides, we are informed, the assumption of an implied promise to act promptly "ignores actuality". "If a court should hold that a contract to decide expeditiously on the proposal did exist, it is believed that, within a short time, all insurance companies doing business in that jurisdiction would incorporate in their applications stipulations expressly negativing any such promise."[21]

And yet, although most courts subscribe to this doctrine, the majority still allows recovery by the back door, so to speak. They regard recovery *ex contractu* as impossible, but at the same time allow recovery *ex delictu*. The failure of an insurance company to take prompt action—according to these decisions—amounts to the breach of a general duty towards the public to act without undue delay on applications for acceptable risks.[22] The courts are sure that the policy of insurance cannot be treated like any other contract. The state, by granting a franchise to the insurance company, by regulating and supervising its business, recognizes the great social importance of insurance business; it is, therefore, in the public interest that applications for acceptable risks shall not be unduly delayed. Thus the courts pay merely lip service to the dogma that the common law of contracts governs insurance contracts. With the help of the law of torts they nullify those parts of the law of contracts which in the public interest are regarded as inapplicable. Disguised as tort law the courts recognized a liability for *culpa in contrahendo* thus making new law with regard to the formation of insurance contracts. This approach enables them to disregard the clause in the application by means of which the company attempted to avoid liability prior to the delivery of the policy. No wonder that this line of reasoning has been sharply criticized not only for its inconsistency[23] but also for undermining legal certainty and the stability of the insurance business. To impose upon an insurance company, because it acts under franchise from the state, a duty to act promptly on an application

20. Prosser, *supra* note 4, at 40, 41. For an analysis of the function of the clause and the efforts of courts to protect the interests of the applicant by manipulating the symbol "delivery," see Patterson, *supra* note 11, at 218, 221, 222.

21. Funk, *The Duty of an Insurer to Act Promptly on Applications* (1927) 75 U. OF PA. L. REV. 207, 214.

22. Duffie v. Bankers' Life Ass'n, 160 Iowa 19, 139 N. W. 1087 (1913); for further cases, Prosser, *supra* note 4, at 41 *et seq.*

23. Savage v. Prudential Life Ins. Co., 154 Miss. 89, 121 So. 487 (1929). "To hold that there is no contract, nor breach of a contract, in failing to insure this applicant, or to notify him that he was not insured, and then to hold that a tort arises, is to hold that there was created a legal duty, and to this we cannot subscribe." *Id.* at 489. The situation is unlike that in the early history of contract law where contractual liability was developed with the help of the action on the case. In the insurance cases contract and tort analyses represent conflicting ideologies.

> would be to open a field of legal liability the limits of which we cannot encompass, and which would go far to introduce chaos in the entire business of insurance, indeed, would almost necessarily reach out into the field of other specially chartered corporations occupying not dissimilar relations to the public, as banks, utility companies, and the like. Public interest more requires that stability of the insurance business which is necessary to guard the great body of persons who enter into relations with it for their own protection and that of those dependent upon them, than it does that certain individuals should be saved the loss which may result by adherence to established legal principles.[24]

This line-up of arguments brings out clearly the basic issue with which the courts in the insurance cases are confronted. It is: can the unity of the law of contracts be maintained in the face of the increasing use of contracts of adhesion? The few courts which allow recovery in contract and the many which allow recovery in tort feel more or less clearly that insurance contracts are contracts of adhesion, and try to protect the weaker contracting party against the harshness of the common law and against what they think are abuses of freedom of contract. The courts denying recovery, on the other hand, cling to the belief that an application for insurance is not different from any other offer, and they are convinced that efforts to build up by trial and error a dual system of contract law must inevitably undermine the security function of all law, particularly since courts are ill equipped to decide whether and to what extent an insurance contract has compulsory features.

To be sure, the task of building up a multiple system of contract law is eminently difficult, particularly since courts are not commissions which are able to examine carefully the ramifications of the problem involved, and can see only the narrow aspect of the total problem which comes up for litigation. Equally difficult is the job of determining whether and to what extent a contract, for instance that of insurance, is a contract of adhesion. Still, the predicament to which an applicant for insurance is exposed by an unreasonable delay in handling his application is deserving of more serious consideration than the assertion that in case of unreasonable delay the applicant can withdraw his offer and apply elsewhere. The denial of liability may very well put a premium on inefficiency. It is submitted that in this respect the attitude of the courts which allow the applicant to recover as if he were insured is more realistic, provided the risk was acceptable and the insurance company, in dealing with the application, deviated from its standard pattern of behavior, on which the applicant could reasonably rely.[25] There has been

24. Swentusky v. Prudential Ins. Co., 116 Conn. 526, 532 (1933).

25. Until now decisions allowing recovery in tort have been rather lax in requiring evidence that the delay in handling the application has caused "loss with-

no evidence that the insurance business has been unable to adjust itself to the new law created by the decisions allowing recovery. This is not surprising since deviations from the standard practice in handling applications which result in "loss without insurance" are the exception.[26]

The idea implicit in the cases which allow recovery seems very fruitful indeed. In dealing with standardized contracts courts have to determine what the weaker contracting party could legitimately expect by way of services according to the enterpriser's "calling", and to what extent the stronger party disappointed reasonable expectations based on the typical life situation.[27] It can hardly be objected that the resulting task of rewriting, if necessary, the contents of a contract of adhesion is foreign to the function of common law courts; the judge-made law in the field of constructive conditions is amply proving the opposite and refutes the contention that a contract implied in fact does not differ from an express contract except that the intention of the party is circumstantially proved.[28]

The task of adjusting in each individual case the common law of contracts to contracts of adhesion has to be faced squarely and not indirectly. This is possible only if courts become fully aware of their emotional attitude with regard to freedom of contract. Here lies the main obstacle to progress, particularly since courts have an understandable tendency to avoid this crucial issue by way of rationalizations. They prefer to convince themselves and the community that legal certainty and "sound principles" of contract law should not be sacrificed to dictates of justice or social desirability. Such discussions are hardly profitable.

To be sure, "case law and the feeling of justice are certainly not synonymous";[29] it is just to obey laws of which one does not approve.

out insurance." For a decision insisting that the plaintiff can recover only if he had reason to believe that a policy would be issued to him and if he were precluded to his damage from procuring other insurance, see Wallace v. Metropolitan Life Ins. Co., 212 Wis. 346, 248 N. W. 435 (1933).

26. The practice of antedating insurance by using premium receipts is a step in the right direction but hardly goes far enough. The applicant, if he is protected at all against the risk of delay, is only protected if he has paid the first premium in full. He is not protected for instance if he has only made a down payment on the first premium, as illustrated by Swentusky v. Prudential Ins. Co., 116 Conn. 526, cited *supra* note 17. Whether and to what extent the applicant is protected depends further on the type of the premium receipt used. For a description of the various types of premium receipts used and their shortcomings, see Comment (1935) 44 YALE L. J. 1223.

27. Llewellyn, Review (1939) 52 HARV. L. REV. 700, at 704. Llewellyn calls our attention to the case law on oral contracts to insure which indeed shows the ingenuity of common law courts in making contracts for the parties on the basis of the typical life situation. See, for instance, Aetna Ins. Co. of Hartford, Conn. v. Licking Valley Millings Co. (C. C. A. 6th, 1927) 19 Fed.(2d) 177; PATTERSON, *supra* note 9 at 59 ff, 64 ff.

28. Prosser, *supra* note 4, at 49.

29. M. Cohen, *Positivism and the Limits of Idealism in the Law*, PROCEEDINGS OF THE SIXTH INTERNATIONAL CONGRESS OF PHILOSOPHY (1927) 469, 470.

But it is equally true that the rules of the common law are flexible enough to enable courts to listen to their sense of justice and to the sense of justice of the community. Just as freedom of contract gives individual contracting parties all the needed leeway for shaping the law of contract according to their needs, the elasticity of the common law, with rule and counter-rule constantly competing, makes it possible for courts to follow the dictates of "social desirability". Whatever one may think about the possibility of separating the "law that is" from the "law that ought to be", this much is certain: In the development of the common law the ideal tends constantly to become the practice. And in this process the ideal of certainty has constantly to be weighed against the social desirability of change, and very often legal certainty has to be sacrificed to progress. The inconsistencies and contradictions within the legal system resulting from the uneven growth of the law and from conflicting ideologies are inevitable.[30]

It is not even profitable to spend "the energy of counsel, the money of clients and the time and analysis of judges"[31] in discussing the problems presented by contracts of adhesion in terms of established legal principles and to proclaim that recovery is "contrary to the well settled principles of contract law". This approach tries to create the impression that the rules concerning the formation of contracts are a closed and harmonious system. But this is hardly the case. The doctrine of consideration, for instance, more than any other doctrine, is in a constant process of evolution, full of contradictions and inconsistencies.[32] It has responded to the belief in freedom of contract, as the pepper corn theory of consideration illustrates. It can also be used to protect a creditor against the risk of economic duress of his debtor (*Foakes v. Beer*).[33] Diametrically opposed social policies have thus been defended in the name of consideration. Furthermore, the harshness of the rule of *Thorne v. Deas*,[34] which

30. Still, in a time of transition like ours where the dichotomy between political freedom and economic insecurity belongs to the experience of everyday life and where the widespread feeling of economic insecurity threatens even political freedom the psychological urge to rely on the law for certainty is particularly powerful. For a penetrating psychological analysis of the emotional dilemma of modern man, see FROMM, ESCAPE FROM FREEDOM (1942).

31. Douglas, *Vicarious Liability and Administration of Risk, I* (1929) 38 YALE L. J. 584, 594.

32. Corbin, *Recent Developments in the Law of Contracts* (1937) 50 HARV. L. REV. 449, 453 *et seq.*

33. L. R. 9 App. Cas. 605 (1884); Sharp, *Promissory Liability, II* (1940) 7 U. OF CHI. L. REV. 250, 253. Thus interpreted the rule in *Foakes v. Beer* is not as conceptualistic as its critics contend.

34. 4 Johns. 84 (N. Y. 1809). It is difficult to believe that the efforts on the part of some courts [*e.g.*, Comfort v. McCorkle, 149 Misc. 826, 268 N. Y. Supp. 192 (1933)] to reconcile §§ 45 and 90 of the Restatement by limiting the application of promissory estoppel to charitable subscriptions and promises to make gifts (non-commercial cases) will be successful in the long run.

seems to support the theory of the courts which deny liability in contract, is mitigated by a counter rule which is constantly gaining in strength and has found expression in Section 90 of the Restatement. Even the mere risk of reliance has been regarded sufficient consideration,[35] a doctrine which comes in handy here to offset the argument that the applicant could have withdrawn his application and applied for insurance elsewhere. It is true that acceptance of the application can hardly be inferred from silence for an unreasonable length of time since the standard clause in the application expressly warns the applicant that the company shall incur no liability under the application until it has been approved and a formal policy has been issued and delivered. But is this clause sufficiently unequivocal to negative not only an acceptance by silence but also an implied collateral promise (as it is technically called) to take prompt action on an application for an acceptable risk?[36] More serious is the argument that the assumption of an implied promise to act promptly is unrealistic because insurance companies, once subjected to such an implied promise, would immediately negative it by express stipulation in the policy. But is this argument not begging the question? The crucial problem is not whether insurance companies would insert such a clause but whether they could do so with impunity.

Thus, technical doctrines of the law of contracts cannot possibly provide the courts with the right answers. They convince only those courts which are already convinced. For instance, which consideration doctrine the court is going to choose as the correct one depends upon its attitude with regard to freedom of contract. All the technical doctrines resorted to by the courts in the insurance cases denying liability are in the last analysis but rationalizations of the court's emotional desire to preserve freedom of contract. Even the cases which hold the insurance company liable in tort pay tribute to the dogma; otherwise it would have been unnecessary constantly to emphasize that the plaintiff is not seeking recovery in contract. The freedom of contract dogma is the real hero or villain in the drama of the insurance cases, but it prefers to remain in the safety of the background if possible, leaving the actual fighting to consideration and to the host of other satellites—all of which is very often confusion to the audience which vaguely senses the unreality of the atmosphere.

35. For a collection and discussion of authorities see Note, *Promissory Obligations Based on Past Benefits or Other Moral Obligations* (1939) 7 U. OF CHI. L. REV. 124, 133 *et seq.*

36. The risk of reliance furnishes the needed consideration. The situation is in a way the reverse to that presented by Los Angeles Traction Co. v. Wilshire, 135 Cal. 654, 67 Pac. 1086 (1902) and similar cases envisaged by § 45 of the RESTATEMENT OF CONTRACTS.

Still, the tort cases are a constant though indirect challenge to the claims of the freedom of contract dogma. They keep alive the question whether or not the "received ideas"[37] on freedom of contract which form the background of the insurance cases represent a cultural lag.

The individualism of our rules of contract law, of which freedom of contract is the most powerful symbol, is closely tied up with the ethics of free enterprise capitalism and the ideals of justice of a mobile society of small enterprisers, individual merchants and independent craftsmen. This society believed that individual and cooperative action left unrestrained in family, church and market would not lessen the freedom and dignity of man but would secure the highest possible social justice. It was firmly convinced of a natural law according to which the individual serving his own interest was also serving the interest of the community. Profits can be earned only by supplying consumable commodities. Freedom of competition will prevent profits from rising unduly. The play of the market if left to itself must therefore maximize net satisfactions. Justice within this framework has a very definite meaning. It means freedom of property and of contract, of profit making and of trade.[38] Freedom of contract thus receives its moral justification. The "prestabilized harmony" of a social system based on freedom of enterprise and perfect competition sees to it that the "private autonomy" of contracting parties will be kept within bounds and will work out to the benefit of the whole.

With the decline of the free enterprise system due to the innate trend of competitive capitalism towards monopoly, the meaning of contract has changed radically. Society, when granting freedom of contract, does not guarantee that all members of the community will be able to make use of it to the same extent. On the contrary, the law, by protecting the unequal distribution of property, does nothing to prevent freedom of contract from becoming a one-sided privilege. Society, by proclaiming freedom of contract, guarantees that it will not interfere with the exercise of power by contract. Freedom of contract enables enterprisers to legislate by contract and, what is even more important, to legislate in a substantially authoritarian manner without using the appearance of authoritarian forms. Standard contracts in particular could thus become effective instruments in the hands of powerful industrial and commercial overlords enabling them to impose a new feudal order of their own making upon a vast host of vassals.[39] This spectacle is all the more fascinating since not more than a hundred years ago contract ideology had been success-

37. Pound, *The New Feudalism* (1930) 16 A. B. A. J. 553, 554.

38. See Hamilton, *Competition*, 2 Encyc. Soc. Sci. 141, 142.

39. See Note, *"Mutuality" in Exclusive Sales Agency Agreements* (1931) 31 Columbia Law Rev. 830 *et seq.*

fully used to break down the last vestiges of a patriarchal and benevolent feudal order in the field of master and servant (*Priestley v. Fowler*). Thus the return back from contract to status which we experience today was greatly facilitated by the fact that the belief in freedom of contract has remained one of the firmest axioms in the whole fabric of the social philosophy of our culture.

The role played by contract in the destruction of the institutional framework of capitalistic society is constantly obscured to the lawyer by the still prevailing philosophy of law which neglects to treat contract as the most important source of law. According to conventional theory contract is only a convenient label for a number of "operative facts" which have the consequences intended by the parties if the law so ordains.[40] In this respect the great philosophers of natural law thought quite differently: society, in proclaiming freedom of contract—according to their teaching—has delegated to individual citizens a piece of sovereignty which enables them to participate constantly in the law making process. Freedom of contract means that the state has no monopoly in the creation of law. The consent of contracting parties creates law also. The law-making process is decentralized. As a result, law is not an order imposed by the state from above upon its citizens; it is rather an order created from below. This was a realistic insight.[41] Unwarranted, however, was the optimistic belief that capitalism meant a permanent advance over the preceding social system, feudalism, because of the fact that contract and not status had become the chief means of social integration. Nor can we subscribe to the thesis of natural law philosophers that the progress in any society towards freedom is to be measured by the extent to which all political relations can be reduced to contract, "the perfect form of obligation".

In the happy days of free enterprise capitalism the belief that contracting is law making had largely emotional importance. Law making by contract was no threat to the harmony of the democratic system. On the contray it reaffirmed it. The courts, therefore, representing the community as a whole, could remain neutral in the name of freedom of contract. The deterioration of the social order into the pluralistic society of our days with its powerful pressure groups was needed to make the wisdom of the contract theory of the natural law philosophers meaningful to us. The prevailing dogma, on the other hand, insisting that contract is *only* a set of operative facts, helps to preserve the illusion that the "law" will protect the public against any abuse of freedom of contract.

40. 1 WILLISTON, CONTRACTS (rev. ed. 1936) § 1; RESTATEMENT, CONTRACTS (1932) § 1; Pound, *Contract*, 2 ENCYC. SOC. SCI. 323.

41. See M. Cohen, *The Basis of Contract* (1933) 46 HARV. L. REV. 553, 585.

This will not be the case so long as we fail to realize that freedom of contract must mean different things for different types of contracts. Its meaning must change with the social importance of the type of contract and with the degree of monopoly enjoyed by the author of the standardized contract.

MICHIGAN LAW REVIEW

Vol. 45 JANUARY, 1947 No. 3

ECONOMIC DURESS—AN ESSAY IN PERSPECTIVE*

John P. Dawson†

THE boundaries of common law duress have been gradually expanding for more than a century. The processes of expansion are themselves of interest, as illustrating methods of growth in a system of case law. More important is the goal toward which this movement aims. For it is through duress and related ideas that private law has dealt most directly with problems raised by inequality in bargaining power. Particularly in the field now known as economic duress, courts have been compelled to take a stand on that central issue of modern politics, the control of economic power. Both the growth in remedial doctrines and the limits that have been set are significant, for in both can be found the trace of broader objectives which have been only partly defined.

Common law doctrines of duress were ill-equipped for the work they were later to do. Progress came first through concentrating on specific type-situations, with growth outward from them. In the next stage the lines became blurred between common law duress and remedial principles in other fields. In the process, some larger ideas filtered

† Professor of Law, University of Michigan.

* This article was prepared in its present form in May, 1942. The only substantial changes made have been in the conclusion, in a few footnotes and through the addition to footnotes of more recent decisions. The whole subject of Compulsory Contracts has been reviewed in the interval in the series of articles appearing in the July, 1943 issue of the COLUMBIA LAW REVIEW. Much of the material appearing in the similar series on Consideration in the May, 1941 issue of the COLUMBIA LAW REVIEW also has an obvious bearing on the subject here considered. The article of Milton Green, "Proof of Mental Incompetency and the Unexpressed Major Premise," 53 YALE L. J. 271 (1944) and the articles on "Economic Duress" by John Dalzell, 20 N.C. L. REV. 237 and 341 (1942) throw much light on important aspects of the general problem. In spite of the extensive literature that has appeared since this article was prepared, it has been decided to publish it in the hope that it will contribute some further information on a subject now rendered familiar.

Any writer in this field owes a great debt to Professor R. L. Hale, whose original article, "Coercion and Distribution in a Supposedly Non-coercive State," 38 POL. SCI. Q. 470 (1923), opened up wide vistas, and who has since published a series of illuminating statements of the same basic point of view. The present article draws heavily on his analysis, and it is a pleasure to record the writer's great indebtedness to him.

in, adding new content and opening new avenues for growth. The question should now be asked whether the older common law doctrines of duress have not been overloaded; or in other words whether they can do their work without a basic revision in which the elements newly added receive their proper emphasis.

In the present survey attention will be concentrated exclusively on the functions of duress and associated doctrines in the area of private contract law. The remedies mainly considered are the group which aim at restitution. Other means of attack on the problem of economic pressure must be postponed to a later stage.

I

COMMON LAW DURESS

The concept of duress which first appeared in common law sources was merely a by-product of legal controls over crime and tort. In Bracton's treatise the specific content given duress was physical assault, exerted or threatened, by means of which transfers (in particular, transfers of land) were extorted.[1] The thirteenth century cases collected in *Bracton's Notebook* likewise involved either physical imprisonment or threats of serious bodily harm.[2] The cancellation of transfers induced by such means was a natural supplement to the sanctions then being evolved for the control of private violence.

The content thus given to doctrines of duress has lasted late. Through the Year-Book period, when duress was mentioned frequently in lawyers' discussions of voidable transactions, discussion was focussed on the case of false imprisonment, with threats of physical harm on the outer fringe.[3] In one instance in the Year-Books duress

[1] BRACTON, DE LEGIBUS, fol. 16b-17. This appears from the illustrations given in Bracton's text; he also states explicitly that the threat "must include the peril of death or bodily torture."

[2] BRACTON'S NOTEBOOK, cases 182, 200, 229, 750, 1126, 1643, and 1913. Strangely enough, in none of these cases is relief for duress actually awarded. The reason is the requirement stated by Bracton himself [DE LEGIBUS, fol. 16b] that disaffirmance must be prompt and that the hue and cry must be raised and the "violence" shown to the king's coroner and later to the county court. The litigants in most of the cases referred to failed to comply with this requirement. The imposition of this requirement further suggests the intimate connection between duress and early criminal law enforcement.

[3] Y.B. 8 H. 6, 7; 2 Edw. 4, 21, pl. 16; 18 Edw. 4, 29, pl. 27; 15 Edw. 4, 1, pl. 2; 21 Edw. 4, 12, pl. 4 and 27, pl. 22; 1 H. 7, Pasch., pl. 2. A plea of duress of imprisonment was held good in an action of debt on a bond in Y.B. 4 Edw.4, 17, pl. 32. See also Y.B. 1 and 2 Edw. 2 (Selden Society Pub., vol. 17), p. 35 (1308); and 13 Rich. 2 (Ames Foundation) 59 (1389).

was expanded to include tortious interference with chattel ownership.[4] But the influence of Coke confirmed the ancient restriction, which remained official doctrine at least to the time of Blackstone.[5]

Even harder to eliminate was another restrictive formula, the requirement that the pressure or threat must be sufficient to overcome a "constant" man. The standard of "ordinary firmness" was borrowed by Bracton from the glossators, through whom it was likewise transmitted to the legal systems of continental Europe.[6] It reappeared in Coke and Blackstone.[7] In American cases of the nineteenth century this formula was frequently reproduced; even in the twentieth century it has occasionally put in an appearance.[8] That it directly influenced decision at any time is unlikely and for present purposes it can be henceforth ignored. Its chief effect was to preserve emphasis on the misconduct of the coercing party, thus distracting attention from the specific consequences to the party coerced..

The extension of duress into the field of economic pressure began in the eighteenth century. The situation which inspired this development involved a relatively simple and clear-cut type of oppression—the wrongful seizure or detention of personal property. The choice was consistent with earlier assumptions as to the central function of duress doctrines. The conduct producing pressure was not merely a threatened, but an actual invasion of chattel ownership independently wrongful by the common law of tort. Restitution of payments made or property transferred under such pressure provided a sanction which merely supplemented the damage remedy already available. Inconvenience to the

[4] 20 Edw. 3 (Lib. Assn.) 72, pl. 14, involving a release of a cause of action for disseisin, the release being held ineffective because executed under the pressure of seizure of the releasor's "beasts." In Y.B. 7 Edw. 4, 21, pl. 24, defendant in an action of debt on a bond sought to avoid his obligation by pleading duress, consisting of a tortious seizure and detention of defendant's "beasts." Counsel for plaintiff argued at length that threats of merely economic injury could not constitute duress. Eminent counsel for defendant, among them Littleton, urged that the plea was sufficient, Littleton being quoted for the statement that it was "all the same to a man to be beaten and to lose all his goods or having his buildings destroyed." The case was unfortunately adjourned and no decision is recorded.

[5] COKE, SECOND INSTITUTE 483 (1642); COKE ON LITTLETON 253b (1633). Similarly, SHEPPARD'S TOUCHSTONE 60-61 (1648); I BLACKS., COMM., Chitty ed., 131 (1859).

[6] BRACTON, DE LEGIBUS, fol. 16b. The survival of the "constant man" as the standard for duress is discussed in relation to French and German law by Dawson, "Economic Duress and the Fair Exchange in French and German Law," 11 TULANE L. REV. 345 at 347-350 (1937).

[7] COKE ON LITTLETON 253b (1633); I BLACKS., COMM., Chitty ed., 131 (1859).

[8] 5 WILLISTON, CONTRACTS, rev. ed., § 1605 (1937).

owner of the goods detained could normally be anticipated, so that close scrutiny of his special position at the time could be dispensed with. Beginning as early as 1732 with *Astley v. Reynolds*,[9] a series of English cases established the common law doctrine of "duress of goods."[9] Relatively narrow, but within its limits distinctly favorable to relief, the law of "duress of goods" provides the starting point, the central type-case, of economic duress.

American courts in the nineteenth century followed the lead of the English cases. Both in actions for restitution of money paid and by defense to the enforcement of contract liability, a series of decisions confirmed this extension of duress into the field of economic pressure.[10] It was at first assumed, as in the English cases, that the wrongfulness of the means used made unnecessary any inquiry into their precise effects on the party coerced. But the need was soon felt for explaining the functions of duress doctrines in terms of a broader objective. The objective defined, for duress of goods cases as in other types of duress, was that of ensuring the freedom of the individual will. The introduction of this idea complicated the clear and simple pattern cut by the early cases. It became necessary to demonstrate not merely that the means used were wrongful, by the tests of the criminal law or the law of tort, but also that the result was substantial interference with freedom of choice, at the particular time and place.

[9] Astley v. Reynolds, 2 Strange 915 (1732), involving a pledge of some plate which defendant, the pledgee, refused to redeliver unless paid an excess over the legal rate of interest on the loan secured by the pledge. Allowing recovery of the excess interest paid by plaintiff in order to secure delivery of the plate, the court said at p. 916: "The plaintiff might have such an immediate want of his goods, that an action of trover would not do his business; where the rule *volenti non fit injuria* is applied, it must be where the party had his freedom of exercising his will, which this man had not: we must take it he paid his money relying on his legal remedy to get it back again."

Later English cases confirming the doctrine of Astley v. Reynolds are Irving v. Wilson, 4 Term Rep. 485 (1791); —— v. Pigott, referred to by Lord Kenyon in 2 Esp. 723 (1799); Shaw v. Woodcock, 7 Barnw. & C. 73 (1827); Hills v. Street, 5 Bing. 37 (1828); Pratt v. Vizard, 5 Barnw. & Ad. 808 (1833); Ashmole v. Wainwright, 2 Q.B. 837 (1841); Oates v. Hudson, 6 Exch. 346 (1851); Green v. Duckett, 11 Q.B. Div. 275 (1883). Included in the category of "goods" were such diverse objects as wagon-loads of hams, life insurance policies, deeds of real estate, and a bull. In none of these cases was any showing required that the wrongful detention produced especially stringent effects on the owner. In Shaw v. Woodcock, supra at 85, Holroyd, J., expressly declared that recovery of the payment made would be allowed "whether there was pressing necessity or not."

[10] Sasportas v. Jennings, 1 Bay (S.C.) 463 (1795); Chamberlain v. Reed, 13 Me. 357 (1836); Quinnett v. Washington, 10 Mo. 53 (1846); Alston v. Durant, 2 Strob. (S.C.) 257 (1847).

The shift in emphasis appeared primarily in the occasional twentieth century decisions that refused to find duress where a damage action or an action of replevin for the goods detained provided an alternative means of relieving the pressure, so that payment of the sum demanded for their release was considered "voluntary."[11] A similar desire to relate "duress of goods" to more generalized theories of duress seemed to be involved in those earlier and more numerous cases which found duress to exist but were quite careful to point out the serious economic effects of a wrongful detention of goods, arising from the special circumstances of the owner and depriving him in fact of effective freedom of choice.[12] It appeared indeed that some courts were headed toward tests phrased in terms of the adequacy of alternative remedies, comparable to those employed in determining the grounds for equity jurisdiction. But on the whole these tests were interpreted liberally. The delays involved in resort to legal remedies were emphasized and the conclusion frequently suggested that an owner of goods should not be compelled to accept their value in money when by paying the sum de-

[11] An action of replevin was held an adequate remedy for securing the personal property detained in Karschner v. Latimer, 108 Neb. 32, 187 N.W. 83 (1922), involving a herd of cattle; and Krouse v. Krouse, 48 Ind. App. 3, 95 N.E. 262 (1911), involving the detention of the clothing of a lawyer by his wife at the time of the San Francisco fire, the court indicating its belief that it was not necessary to be well dressed in order to practice law in San Francisco.

A damage remedy was held adequate protection in K., M., & O. Ry. Co. v. Graham, (Tex. Civ. App. 1912) 145 S.W. 632, and Furman v. Lanahan, 159 Md. 1, 149 A. 465 (1930), the latter case involving, however, the detention of *money* deposited by a customer with a stockbroker. Williams v. Phelps, 16 Wis. 83 (1862), and Burgess v. Commercial Nat. Bank, 144 Wis. 59, 128 N.W. 436 (1910), both employ a much stricter test than is usually imposed.

[12] Harmony v. Bingham, 12 N.Y. 99 (1854), goods "of great value" intended for sale in Mexico and detained in Missouri for freight charges not due, the court concluding that plaintiff's necessity for obtaining immediate possession was "urgent"; Beckwith v. Frisbie, 32 Vt. 559 (1860), "pressing necessity" of immediate possession deduced from declining market price of oats detained; Cobb v. Charter, 32 Conn. 358 (1865), emphasis placed on inconvenience to mechanic through detention of chest of tools needed for earning a livelihood. Similarly, Stenton v. Jerome, 54 N.Y. 480 (1873); McPherson v. Cox, 86 N.Y. 472 (1881); Motz v. Mitchell, 91 Pa. 114 (1879); Fargusson v. Winslow, 34 Minn. 384 (1885); Oliphant v. Markham, 79 Tex. 543, 15 S.W. 569 (1891); Exporters & Graders Compress & Warehouse Co. v. Spivey, (Tex. Civ. App. 1923) 249 S.W. 1086; Glass & Co. v. Haygood, 133 Ala. 489 (1901); DuVall v. Norris, 119 Ga. 947, 47 S.E. 212, (1904); Tandy v. Elmore-Cooper Live Stock Commission Co., 113 Mo. App. 409, 87 S.W. 614 (1905); McTigue v. Arctic Ice Cream Supply Co., 20 Cal. App. 708, 130 P. 165 (1912); Carhill Petroleum Co. v. Ennis-Bayard Petroleum Co., 81 Pa. Super. 486 (1923); Jones v. Sherwood Distilling Co., 150 Md. 24, 132 A. 278 (1926).

manded the delivery of the goods themselves could be secured.[13] Even these relaxed tests have been ignored in many modern cases, and duress has been found without specific showing of inconvenience or serious injury through the detention.[14] In short, "duress of goods" has been preserved as a special type of economic pressure, for which relief is readily given, although the search for a broader formula has somewhat complicated prediction.

Meanwhile, toward the middle of the nineteenth century, another line of growth had developed, in cases involving overcharges for service by common carriers. At the outset this new category somewhat overlapped "duress of goods," since the demand for an excessive freight rate by a common carrier was frequently accompanied by refusal to surrender goods already in the carrier's custody.[15] Furthermore, the grant of restitution for any overcharge paid by the owner or consignee seemed to present no serious issues of policy, for the common law had already progressed to the point of admitting a liability in damages for the carrier's refusal of service unless paid a discriminatory rate. The relatively simple transfer of ideas that was involved is illustrated by the first case in this new series, the English case of *Parker v. Great Western Ry. Co.*,[16] decided in 1844. Here the railroad's tort liability in damages for unjustified refusal of service was used by the court as an alternative ground for awarding restitution of payments exacted in excess of its published rates. At the same time, however, the court announced a

[13] Chase v. Dwinal, 7 Greenl. (Me.) 134 (1830); Wilkerson v. Hood, 65 Mo. App. 491 (1896). Similarly, Harmony v. Bingham, 12 N.Y. 99 (1854); Lowenstein v. Bache, 41 Pa. Super. 552 (1910). In Miller v. Eisele, 111 N.J. 268, 168 A. 426 (1933), the inadequacy of alternative remedies against a stockbroker was found to consist chiefly in the possibility of his insolvency, making a judgment for damages uncollectible.

[14] Southwestern Alabama Ry. Co. v. Maddox & Son, 146 Ala. 539, 41 S. 9 (1906); Whitlock Machine Co. v. Holway, 92 Me. 414, 42 A. 799 (1899); J. Abrams & Co., Inc. v. Clark, 298 Mass. 542, 11 N.E. (2d) 449 (1937); Nelson v. Nelson, 99 Neb. 456, 156 N.W. 1036 (1916); Berger v. Bonnell Motor Car Co., (N.J. 1926) 133 A. 778; Meier v. Nightingale, 133 N.J.L. 400, 44 A. (2d) 409 (1945); Baldwin v. Liverpool and Great Western Steamship Co., 74 N.Y. 125 (1878); Richmond v. Union Steamboat Co., 87 N.Y. 240 (1881); Cowley v. Fabien, 204 N.Y. 566, 97 N.E. 458 (1912); Clancy v. Dutton, 129 App. Div. 23, 113 N.Y.S. 124 (1908); Smith v. Houston Nat. Exchange Bank, (Tex. Civ. App. 1918) 202 S.W. 181; Johnson v. Townsend & Co., 161 Wash. 332, 296 P. 1046 (1931); The John Francis, (D.C. Ala. 1911) 184 F. 746.

[15] This appears, for example, in the first English case of this type, where the recovery of the overpayment was explained solely in terms of "duress of goods." Ashmole v. Wainwright, 2 Q.B. 837 (1841). Other examples of this coincidence appear in the railroad cases referred to in note 17.

[16] 7 M. & G. 253 (1844).

much broader ground for enforcing restitution, i.e., the economic disparity between shipper and carrier, which compelled the shipper to acquiesce in the carrier's demands. It was the latter theme that succeeding cases developed. Inequality of bargaining power, the inevitable product of state-conferred monopoly, was used to justify this extension of the doctrine of economic duress. With the expanding industrialism of the later nineteenth century, the need for this new category was clearly felt and made articulate. The economic and political power of the railroads made them at first the focal point of the new doctrine.[17] Soon other types of utilities were included.[18] The developing systems of direct rate control had the effect of confirming rather than restricting this supplementary form of relief, through private law remedies.[19] In

[17] McGregor v. Erie Ry Co., 35 N.J.L. 89 (1871); Lafayette & Indianapolis R.R. Co. v. Pattison, 41 Ind. 312 (1872); Chicago & Alton R.R. Co. v. Chicago, V. & W. Coal Co., 79 Ill. 121 (1875); Mobile & Montgomery Ry. Co. v. Steiner, McGehee & Co., 61 Ala. 559 (1878); Peters, Ricker & Co. v. Railroad Co., 42 Ohio St. 275 (1884); Heiserman v. Burlington C. R. & N. R. R. Co., 63 Iowa 732, 18 N. W. 903 (1884); W. Va. Transportation Co. v. Sweetzer, 25 W. Va. 434 (1885); Galesburg & Great Eastern R. R. Co. v. West, 108 Ill. App. 504 (1903); Southwestern Ala. Ry. Co. v. Maddox & Son, 146 Ala. 539, 41 S. 9 (1906); Pine Tree Lumber Co. v. C., R. I. & P. Ry. Co., 123 La. 583, 49 S. 202 (1909); St. Louis & San Francisco R. R. Co. v. Gorman, 79 Kan. 643, 100 P. 647 (1909); Fairford Lumber Co. v. Tombigbee Valley R. R. Co., 165 Ala. 275, 51 S. 770 (1910); Missouri Pac. R. R. Co. v. Fields Bros., 134 Ark. 273, 203 S.W. 1036 (1918); West Construction Co. v. Seaboard Airline Ry. Co., 141 Tenn. 342, 210 S.W. 78 (1918); California Adjustment Co. v. A., T., & S. Ry. Co., 179 Cal. 140, 175 P. 682 (1918); Bowers v. M., K., & T. Ry. Co., (Tex. Civ. App. 1922) 241 S.W. 509; S. D. Warren Co. v. Maine Central Ry. Co., 126 Me. 23, 135 A. 526 (1926).

[18] Gas: New Orleans Gas Light & Banking Co. v. Paulding, 12 Rob. (La.) 378 (1845); Indiana Natural & Ill. Gas Co. v. Anthony, 26 Ind. App. 307, 58 N.E. 868 (1900); Young v. Brooks, (Mo. App. 1933) 56 S.W. (2d) 794; City of Saginaw v. Consumers Power Co., 304 Mich. 491, 8 N.W. (2d) 149 (1943).

Water: Westlake & Button v. City of St. Louis, 77 Mo. 47 (1882); Panton v. Duluth Gas & Water Co., 50 Minn. 175, 52 N.W. 527 (1892); St. Louis Brewing Assn. v. City of St. Louis, 140 Mo. 419, 37 S.W. 525 (1897); American Brewing Co. v. City of St. Louis, 187 Mo. 367, 86 S. W. 129 (1904); City of Chicago v. N.W. Mutual Life Ins. Co., 218 Ill. 40, 75 N.E. 803 (1905); Green v. Byers, 16 Idaho 178, 101 P. 179 (1909); Gess v. Nampa & Meridian Irr. Dist., 33 Idaho 189, 192 P. 474 (1920); Clough v. Boston & Me. R. R., 77 N.H. 222, 90 A. 863 (1914); Barnes Laundry Co. v. Pittsburgh, 266 Pa. 24, 109 A. 535 (1920); B & B Amusement Enterprises Inc. v. City of Boston, 297 Mass. 307, 8 N.E. (2d) 788 (1937).

Electricity: Manhattan Milling Co. v. Manhattan Gas & Electric Co., 115 Kan. 712, 225 P. 86 (1924); City of Boston v. Edison Electric Ill. Co., 242 Mass. 305, 136 N.E. 113 (1922); Piedmont Power and Light Co. v. L. Banks Holt Mfg. Co., 183 N.C. 327, 111 S.E. 623 (1922).

Also Lehigh Coal & Nav. Co. v. Brown, 100 Pa. 338 (1882).

[19] For example, in Clough v. Boston & Me. R.R., 77 N.H. 222, 90 A. 863 (1914), the decision was rested also on the alternative ground that the statute regu-

these cases, as in "duress of goods," some basis was needed for an inference that economic pressure, existing in fact, supplied the motive for the excessive payment.[20] Ordinarily, however, the mere threat of refusal of service sufficed to prove the existence of pressure, without further evidence that inability to secure the service would cause serious dislocation or inconvenience to the person requesting it.[21]

As a result of these developments, two main types of economic pressure—duress of goods and refusal of service by public utilities—were marked off as deserving special treatment. The insight gained in formulating remedial principles for these two type situations began to affect solutions in related areas. One index of this enlarged perspective was the expansion of the categories themselves. The "goods" whose detention would constitute duress were coming to include much more than horses, corn, or carpenters' tools—not only corporate securities but other documentary evidence essential for the enforcement of legal claims.[22] The mere refusal to co-operate in signing proofs of loss,

lating railroad rates was intended for the protection of individuals dealing with the railroad and therefore by implicatioan authorized restitution of overpayments.

Some complication was introduced, however, where the grant of power to regulatory commissions included power to enforce restitution in favor of users of service who had paid excessive rates. To avoid interference with the regulatory function numerous cases reached the conclusion that the administrative remedy must be exhausted before resort to ordinary judicial remedies. City of Boston v. Edison Electric Ill. Co., 242 Mass. 305, 136 N.E. 113 (1922); Texas & Pacific Ry. Co. v. Abilene Cotton Oil Co., 204 U.S. 426, 27 S. Ct. 350 (1907), and cases discussed in 21 IOWA L. REV. 751 (1936).

[20] In Hardaway v. Southern Ry. Co., 90 S.C. 475, 73 S.E. 1020 (1911), for example, excessive freight paid by the shipper was held not recoverable where the payment was made after delivery of the goods consigned and the evidence did not indicate any other element of dependence by the payor on the railroad's services. In some cases the absence of pressure was deduced from the fact that no specific threat of refusal of service was made by the utility. National Enameling & Stamping Co. v. City of St. Louis, (Mo. 1931) 40 S.W. (2d) 593; Central States Power & Light Co. v. Thompson, 177 Okla. 310, 58 P. (2d) 868 (1936). In these cases the decisive factor seems to have been, however, the long delay and acquiescence in the excessive charges by the user of the service, indicating that economic pressure did not supply the motive for the payment. This factor is more clearly emphasized in Killmer v. N. Y. Central & Hudson River R. R. Co., 100 N.Y. 395, 3 N.E. 293 (1885); Monongahela Nav. Co. v. Wood, 194 Pa. 47, 45 A. 73 (1899) and Illinois Glass Co. v. Chicago Telephone Co., 234 Ill. 535, 85 N.E. 200 (1908). See also Kansas City, M. & O. Ry. Co. v. Graham & Price, (Tex. Civ. App. 1912) 145 S.W. 632.

[21] This is most clearly stated in B & B Amusement Enterprises, Inc. v. City of Boston, 297 Mass. 307, 8 N.E. (2d) 788 (1937), but it forms the underlying assumption of many of the cases cited in notes 17 and 18.

[22] Corporate securities: Stenton v. Jerome, 54 N.Y. 480 (1873); Walz v. Muir, 218 App. Div. 495, 218 N.Y.S. 529 (1926); Lowenstein v. Bache, 41 Pa. Super.

needed for the collection of insurance money due, was seen to involve a misuse of economic power hardly distinguishable from the detention of specific chattels.[23] Similarly, the exactions of agencies not ordinarily classed as public utilities could readily be brought within the policies framed for the utility cases, provided a comparable control of the service or function in question was shown to exist in fact.[24]

Defining the limits of these two categories focussed attention on numerous related situations in which the issues of policy involved were more complex. One form of pressure that came into increasing prominence in nineteenth century cases was the threat of criminal prosecution, used as a means of enforcing civil satisfaction. Here the development of remedial doctrines could be in part explained by describing the means of pressure used as improper, since procedures intended for the vindication of public interests were used for private advantage.[25] This explanation was not available, however, and severe limits on restitution were maintained, in that large group of cases where the means of pressure used was the start, or the threat, of civil action. This explanation was equally inapplicable to economic pressure exerted through the mere withholding of goods or services, where interference with chattel ownership or refusal of service by a "utility" was not involved.

In net result no shift in basic premises had yet been achieved by the end of the nineteenth century. With both "duress of goods" and the

552 (1910); Interurban Construction Co. v. Hayes, 191 Mo. 248, 89 S.W. 927 (1905); Johnson v. Townsend and Co., 161 Wash. 332, 296 P. 1046 (1931).

Various forms of documentary evidence: Oliphant v. Markham, 79 Tex. 543, 15 S.W. 569 (1891); Huggins v. Hill, (Mo. App. 1922) 245 S.W. 1105; Moore v. Putts, 110 Md. 490, 73 A. 149 (1909); Motz v. Mitchell, 91 Pa. 114 (1879).

[23] Guetzkow Bros. v. Breese, 96 Wis. 591, 72 N.W. 45 (1897), where inability to secure the proceeds of a fire insurance policy was shown to have threatened disruption of the going business of the insured. Refusal to co-operate in preparing needed documents was held to constitute duress, with much less of a showing of threatened loss, in Kelley v. Caplice, 23 Kan. 337 (1880).

[24] Niedermeyer v. Curators of University of Missouri, 61 Mo. App. 654 (1895), exaction of excessive tuition fee under threat to refuse admission to University; Buckley v. City of New York, 30 App. Div. 463, 52 N.Y.S. 452 (1898), affd., 159 N.Y. 558, 54 N.E. 1089 (1898), payment of fee exacted by city authorities for permit to construct vault in building under construction; Oceanic Navigation Co. v. Stranahan, 214 U.S. 320, 29 S. Ct. 671 (1909), refusal of port authorities to clear vessel at port unless fine paid. Related to these cases of course are the numerous decisions involving license fees or taxes exacted for the privilege of doing business. These cases, with the rest of the tax cases, are laid aside in the present survey, as involving considerations of policy peculiar to the tax field.

[25] As in Morse v. Woodworth, 155 Mass. 233, 27 N.E. 1010, 29 N.E. 525 (1891), and additional cases cited by 5 WILLISTON, CONTRACTS, rev. ed., § 1611 (1937).

refusal of service by public utilities the means used could still be analyzed as "wrongful." In one case the standards of conduct were derived from the common law of tort, in the other from statute and administrative regulation. Restitution doctrines were still merely a supplement to rules evolved for other purposes. The critical questions would only arise as attention shifted to new situations, in which the means of pressure used were for most purposes permissible. Before these situations are reviewed, however, an attempt should be made to describe the ideas that had meanwhile been filtering in from other directions.

II

UNDUE INFLUENCE IN EQUITY

Equity doctrines of undue influence had been concerned from the outset with a different type of inequality of bargaining power. They were never conceived, like common law doctrines of duress, as a corollary of the law of crime and tort. They were aimed instead at protection for the mentally or physically inadequate, whose inadequacy fell short of a total lack of legal capacity. Protection for such persons did not need to be justified through some violation, accomplished or threatened, of the law of tort or crime; indeed it was seldom that the pressure used would include any element of "wrong" as defined by damage action or criminal prosecution It was enough that the extraction of economic gain from persons mentally or physically handicapped was condemned by prevailing standards of ethics, defined and applied by equity courts through their own independent tests.[26]

The difference in fact content of the cases, and the separation maintained for centuries in the administration of legal and equitable remedies, made possible a vigorous growth of equity doctrines from radically different premises. Throughout the formative period doctrines of undue influence were frequently reinforced by other protective doctrines of equity, particularly those evolved for "confidential" and "fiduciary" relationships. Close family relationships frequently provided oppor-

[26] The early Chancery reports are too brief to reveal in any detail the specific grounds for decision. But in Bac. Rep. by Ritch., p. 33, there is an account of a case of this type, decided by Bacon in 1617 and involving cancellation of inter vivos transfers and also a will made by a decrepit old man through what would now be called undue influence. In Clarkson v. Hanway, 2 P. Wms. 203 (1723), a conveyance made by a "weak man," seventy-two years old who was "easily to be imposed," was set aside, the inadequacy of the consideration being mentioned as an additional ground. In Blake v. Johnson, Prec. in Chan. 142 (1700), and Lucas v. Adams, 2 Mod. Cas. in Law and Equity 118 (1725), similar doctrines were applied.

tunity for the exercise of "influence"; or if the parties were not related by blood or marriage, a condition of dependence by the weaker party might provide the elements of a "confidential" relationship which supplemented undue influence as ground for overhauling the transaction. For present purposes it is unnecessary to review the multitude of decisions establishing special tests for "confidential" or "fiduciary" relations, though they bulk very large in modern law and obviously introduce important qualifications to a regime of "free contract."[27] Of immediate interest in the present discussion are those situations in which judicial review could be rested only on impairment of bargaining power, resulting usually from physical or mental weakness.[28]

It was not until the nineteenth century that serious efforts were made to explain the undue influence cases in terms of a larger objective. The objective chiefly employed soon acquired a remarkable appeal, since it coincided with main movements in nineteenth century thought. The "wrong" involved in undue influence, it was said, was the interference with another's will, which should ideally be free. The test for the existence of undue influence became the presence or absence of "free agency," whether or not the individual will had been "overpowered." From this it was easy to move to the broader thesis that, whatever the means of pressure used, "the inequity of the act consists in compelling a person to do what he does not want to do."[29] The objective defined for cases of mental or physical weakness began to seem equally appropriate for situations included in common law duress, such as threats

[27] The discussion in 41 COL. L. REV. 707 (1941) is excellent.

[28] The following cases are illustrative: Blachford v. Christian, 1 Knapp 73 (1829); Stewart v. Stewart, 30 Ky. 183 (1832); Central Bank of Frederick v. Copeland, 18 Md. 305 (1862); Yount v. Yount, 144 Ind. 133, 43 N.E. 136 (1895); Kellogg v. Kellogg, 21 Colo. 181, 40 P. 358 (1895); Benn v. Pritchett, 163 Mo. 560, 63 S.W. 1103 (1901).

[29] The statement quoted appears in Parmentier v. Pater, 13 Ore. 121 (1885). The case was one in which defendant set up a release of the debt sued on, made by the obligee while extremely disturbed, a few days before he committed suicide. The opinion contains the following passage: "I am not able to discover any difference between wrongful means resorted to in order to compel a party to do something against his will. It is no more wicked, in my opinion, to put a person in prison, or to deprive him of a limb, than to scare him to death by any artifice or chicanery that may be employed; the inequity of the act consists in compelling a person to do what he does not want to do. Any course calculated to excite alarm, which is resorted to by one party in order to coerce another to do an act detrimental to his rights, and advantageous to the former, is unlawful; and I do not think the law should make any distinction between means that are adopted to secure such ends."

Though perhaps more exereme than most, the opinion in Parmentier v. Pater could be matched with many others of the time.

of criminal prosecution and even duress of goods.[30] Inspired by this new conception, the nineteenth century cases seemed to have set off in pursuit of an ideal as attractive as it was unattainable.

Even in the undue influence cases themselves, the ideal of complete freedom for the individual will was incompletely realized. In the first place, it was clear that no legal agencies would entirely eliminate the pressures that operate on the physically, mentally, or emotionally handicapped, or insulate them from all the multiplied stimuli of a complex social environment. The problem, here as elsewhere, is to select the means of pressure that are permissible and to regulate the manner in which they may be exercised. A closer reading of the undue influence cases reveals the operation of some objective tests, side by side with the analysis of individual motives that is chiefly accented in judicial opinions. Transactions must be judged not only in terms of motive but in terms of their effects. For example, in bargain transactions, involving some form of exchange, many cases have revealed a primary emphasis on the "adequacy of the consideration" as a correlative test of the fairness of the whole transaction.[31] Even in gift transactions, which constitute the main bulk of the undue influence cases, objective tests have been used. The aim is by no means to eliminate but to safeguard the powers of donation of the aged, the timid, the physically or mentally weak. Therefore the question, difficult as it is to answer, must be whether existing opportunities for the exercise of pressure have been used to divert the gift from its normal and natural course, in view of the donor's total situation—economic, psychological, and emotional.[32]

[30] A twentieth century illustration is Brown County Bank v. Hage, 156 Minn. 460, 195 N.W. 275 (1923), a case involving threat of criminal prosecution and provoking comments remarkably similar to those quoted in the previous note from Parmentier v. Pater. Of the duress of goods cases cited in the previous section, many propose the removal of interference with freedom of the individual will as the objective of duress doctrines.

[31] Gartside v. Isherwood, 1 Brown's Ch. C. 558 (1783); Marron v. Marron, 19 Cal. App. 326, 125 P. 914 (1912); Hinkley v. Wynkoop, 305 Ill. 115, 137 N.E. 154 (1922); Thatcher v. Kramer, 347 Ill. 601, 180 N.E. 434 (1932); Cole v. Henning, 237 Mich. 108, 211 N.W. 82 (1926).

[32] Since the above passage was written, very similar conclusions have been stated after an extensive review of the cases by Milton Green, "Proof of Mental Incompetency and the Unexpressed Major Premise," 53 YALE L. J. 271 (1944). Though Green's discussion is primarily aimed at the issue of mental competency, the overlapping of mental incompetency and undue influence doctrines in their application and the substantial identity of the basic problems involved make the passages referred to (esp. pp. 298-311) decidedly relevant here. Even more significant is the reluctance of the courts, described by Green and evident throughout the undue influence cases, to

The success of the courts in attaining their declared objective is of less concern at the moment than the enlarged perspective toward the whole problem of coercion that undue influence cases supplied. It was the undue influence cases that helped most to exorcise the ghostly figure of the "constant man," who had stalked the fields of common law duress since the time of Bracton; for in the undue influence cases it was abundantly clear that the weak, the timid, the anxious and submissive were precisely the ones who should and did receive the greatest legal protection. By posing the problem of individual freedom and condemning some of the subtler forms of compulsion, the undue influence cases suggested to many courts a new approach to those other types of pressure, particularly economic pressure, which were being brought within the scope of common law duress.[33]

Toward the end of the nineteenth century the effects of the undue influence cases were increased by the breakdown of procedural distinctions between law and equity. So long as doctrines of undue influence could be confined to "equity" cases, there remained at least a procedural distinction which preserved the purity of common law doctrine. But in the late nineteenth century and increasingly in the twentieth, undue influence was made available in law actions, both by way of defense and by way of affirmative action for restitution.[34]

Accompanying the breakdown of divisions between law and equity there came an expansion of the content of undue influence to include various cases of psychological pressure without extreme disparities in mental or physical condition of the parties. Precise definition of the elements of undue influence had never been undertaken, indeed, had been carefully avoided. But it was clear that the doctrine was being extended to numerous peripheral situations, just as common law duress was being pushed out beyond the typical cases on which attention had been mainly focussed. This widening range of application obscured still

disclose their "unexpressed major premise"—that the *motives* in such transactions cannot be judged apart from their *results*.

[33] Perhaps the most influential and the most widely quoted of the later cases was Galusha v. Sherman, 105 Wis. 263, 81 N.W. 495 (1900), in which Judge Marshall was chiefly concerned, in his extensive review of the cases, with rejecting the standard of reasonable firmness. The whole course of the argument, however, reflected the important developments that had by then occurred in "modern" doctrines of duress.

[34] Illustrative cases are cited by 5 WILLISTON, CONTRACTS, rev. ed., § 1625 (1937). To them should be added Woodbury v. Woodbury, 141 Mass. 329, 5 N.E. 275 (1886), and Eldridge v. May, 129 Me. 112, 150 A. 378 (1930), both of which allow quasi-contract restitution of money paid through undue influence; also Parmentier v. Pater, 13 Ore. 121 (1885).

further the boundaries between undue influence and other types of coercion.[85]

More rapid growth was prevented, however, by a basic contradiction in the concepts of "freedom" which were now at work. On the one hand, doctrines of undue influence were attempting to "free" the individual by regulating the pressures that restricted individual choice; on the other hand, theories of economic individualism aimed at an entirely different kind of freedom, a freedom of the "market" from external regulation. It was not yet fully recognized that the freedom of the "market" was essentially a freedom of individuals and groups to coerce one another, with the power to coerce reinforced by agencies of the state itself. Even though the larger implications of this idea were by no means understood, one simple and quite obvious deduction had already been made—that is, that if the "market" was to be free, any form of external regulation was objectionable. Regulation by court-enforced rules of private law seemed just as unwise and dangerous as regulation by statute or administrative action. From this point of view, where urgent need or special disadvantage compelled agreement to the terms proposed, these circumstances must be disregarded since they differed only in degree from the basic conditions which governed the exchange of goods and services throughout society.

To resolve this major dilemma more perspective was needed than was supplied by the materials of individual cases or the techniques of refined distinction in a system of case law. Indeed it appears that the insight gained through a close analysis of the undue influence cases helped to postpone the needed perspective, through directing attention toward a false issue. Lacking a general theory to explain the results desired, the courts approached the undue influence problem through analysis of the "will" which undue influence destroyed. It is true that in some of the more extreme cases the condition of the person "unduly" influenced might almost be described as one of complete subjection, with a "substitution" of the will of another. But even in the more extreme cases, it was usually inaccurate to say that the transfer or agreement did not result from an exercise of volition, however narrow the

[85] Schoellhamer v. Rometsch, 26 Ore. 394, 38 P. 344 (1894), high pressure salesmanship used on immigrant woman physically and nervously exhausted by her trip; Heckman v. Heckman, 215 Pa. 203, 64 A. 425 (1906), husband's refusal to live with wife, coupled with "persistent importunities"; Kocourek v. Marak, 54 Tex. 201 (1881), threat of husband to abandon wife; Faulkner v. Faulkner, 162 App. Div. 848, 147 N.Y.S. 745 (1914), threat of wife to abandon husband, plus "vile names"; Trigg v. Trigg, 37 N.M. 296, 22 P. (2d) 119 (1933), threat of wife to abandon husband, plus "nagging"; Kuelkamp v. Hidding, 31 Wis. 503 (1872).

range within which volition could operate. In the cases of less extreme disparity, the search for evidence that the "will" had been destroyed confused the reasoning of the undue influence cases. More important than this, it left a legacy in the broader field of duress and helped to confirm the impression that relief for any type of pressure depends on a showing of complete absence of consent. This impression had filtered down from the cases of threats of physical violence, the original source of duress doctrines. Even in this type of case, courts had been slow to realize that the instances of more extreme pressure were precisely those in which the consent expressed was *more* real; the more unpleasant the alternative, the more real the consent to a course which would avoid it.[26]

It can be concluded, then, that the effect of the undue influence cases was an important extension of remedial doctrines and the development of some new insight into the broader problems of coercion. But the insight was only partial, since the theory which guided it was incomplete. Through excessive emphasis on one element—the impaired "will" of the coerced party—attention was diverted from other crucial elements. The whole process of growth and coalescence produced not clarification but confusion, since an organizing principle was not yet clearly seen.

III

PROTECTION OF EXPECTANT HEIRS IN EQUITY

In the meantime another body of doctrine had taken shape out of the haze which surrounds the earlier activities of the English Chancery. This doctrine, giving relief to expectant heirs, promised at first no major contribution. It aimed at a narrow objective, the protection of a landed aristocracy against its own improvidence.

[26] This point is made particularly by Dalzell, "Duress by Economic Pressure," 20 N.C. L. REV. 237 at 239-40 (1942), and by Hale, "Bargaining, Duress and Economic Liberty," 43 COL. L. REV. 603 at 616-17 (1943). The latter writer proceeds to point out the irrelevancies that are injected into judicial opinions by the search for evidence of emotional disturbance or dismay in the person subjected to threat.

The notion that compulsion negates consent arises of course through importing into the concept of "consent" a whole body of assumptions as to the degree of freedom of choice that normally characterizes the relations of the individual to his environment. Once these assumptions are rejected and the universality of pressure restricting choice has been recognized, it becomes clear, as Patterson has said, that "The attempt to solve legal problems by the touchstone of 'free will,' by postulating an individual will insulated from its social environment, only serves to obscure the genuine problems of ethics and policy. If a man is deemed not to 'consent' because he was induced by pressure outside himself, then consent becomes a useless concept in the administration of justice." Patterson, "Compulsory Contracts in the Crystal Ball," 43 COL. L. REV. 731 at 741 (1943).

The need for special protection arose from prevailing standards of fashionable conduct, which dictated a high level of conspicuous expenditure. It was honorable to maintain a gentlemanly extravagance, even where current income did not suffice: all blame was reserved for those hangers-on of polite society who pandered to extravagance by supplying it with the necessary means. Their special prey was the expectant heir or reversioner, in urgent need of ready cash but precluded from borrowing at moderate rates of interest by the uncertainty of his prospects. The Chancellors of the late seventeenth and eighteenth centuries imposed a firm control on traffic with such persons. In the Chancery reports they published their determination to prevent dispersal of family estates through improvident sales by their prospective owners. The motive was clear—to preserve for a dominant class the economic resources on which its prestige and power depended.[87]

The audience for which the Chancery reports were intended did not expect apologies for the policies so announced. Nevertheless the need was felt for some broader ground than the protection of class interest. The broader ground adopted involved a combination of two elements—the economic necessity which impaired the bargaining power of the expectant heir and the unequal exchange of values which this necessity produced. Both elements had implications extending far beyond the immediate problem. Even within the radius of the immediate problem, the refinement and elaboration of these elements produced an important movement of ideas.

From an early stage the class of "expectant heirs" was conceived

[87] Lord Nottingham is as emphatic a spokesman as any. In Berney's Case (1680), quoted from Nottingham's manuscripts in 2 Swanst. 142, 143, relief was given against stipulations for penalties and against sales of goods to a necessitous heir at excessive prices, on repayment by the heir of the sums of money actually borrowed. The reasons stated for allowing relief were that "this infamous kind of trade and circumvention ought by all means to be suppressed; the Star Chamber used to punish it, and this Court did always relieve against it. No family can be safe if this be suffered."

Again, in Cole v. Gibbons, 3 P. Williams 290 at 293 (1734), Lord Talbot is quoted to the effect that it is "the policy of the nation to prevent what was a growing mischief to ancient families, that of seducing an heir apparent from a dependence on his ancestor who probably would have supported him, and, by feeding his extravagancies, tempting him in his father's life-time, to sell the reversion of that estate, which was settled upon him; forasmuch as this tended to the manifest ruin of families. . . ."

The snobbery with which these cases reek is well conveyed by Earl of Portmore v. Taylor, 4 Sim. 182 at 213 (1831): "The mere fact, that Lord Portmore was not only the Heir Apparent of his Father, but also the expectant Heir to a Peerage, also brings the Case distinctly within the Rule laid down by this Court, and which is founded on general policy, namely, that this Court will not allow the Heir of a Family of Rank to be reduced to poverty and distress by dealing with his expectancies."

broadly, to include remaindermen, reversioners, and other persons temporarily distressed who had hopes of future prosperity. Transfers of expectancies, as such, appeared very rarely, though where they did the need for employing the machinery of specific performance of contract made available the higher standards of fairness evolved in specific performance remedies.[38] The English cases centered chiefly around transfers of reversions or remainders, in return for present cash.[39] Relief was also extended to the case of the so-called *post obit* bond, by which a loan of money was secured through a promise to pay a much larger sum on the death of a person or persons with outstanding life or entail interests.[40] In the earlier cases, the Chancellor's vigilance likewise reached out to transactions, such as sales of goods at exorbitant prices or grants of rent-charges or annuities, which seemed to disguise mere loans of money and involved an unequal exchange.[41]

As in more modern regulations of the small-loan market, these doctrines encountered difficulties. Even to contemporaries it was evident that too restrictive a control by equity might prove harmful to the very persons whom the Chancellor aimed to protect, by depriving them of

[38] Resort to doctrines of specific performance was of course made necessary by the refusal of common law courts to recognize "property" in mere expectancies, which did not fit within the categories of reversion or vested remainder. I SIMES, FUTURE INTERESTS, § 234 (1936). Since enforcement depended on contract doctrines, it was possible to employ the usual requirements of fairness and adequacy of price developed for the remedy of specific performance. Cases allowing specific performance, subject to such requirements, are Beckley v. Newland, 2 P. Williams 182 (1723); Hobson v. Trevor, 2 P. Williams 191 (1723); Medcalfe v. Ives, 1 Atk. 63 (1737); Whitfield v. Fausset, 1 Ves. Sr. 387 (1750); Wright v. Wright, 1 Ves. Sr. 409 (1750); Wethered v. Wethered, 2 Sim. 183 (1828).

[39] Cancellation was ordered on repayment of the sum advanced plus simple interest, in Nott v. Hill, 1 Vern. 167 (1683), transfer of remainder in tail of land worth £800 but subject to outstanding life estate, in return for £30 cash and £20 annuity; Twistleton v. Griffith, 1 P. Williams 310 (1716), sale for £1050 of an estate worth £150 a year but subject to the life estate of plaintiff's aged and infirm father who died within two years; Barnardiston v. Lingood, 2 Atk. 133 (1740), contract to transfer remainder in tail in land worth £300 a year, in return for £300 in cash; Crowe v. Ballard, 1 Ves. 214 (1790), sale for £310 of £1000 legacy payable on death of a woman aged sixty-nine; Peacock v. Evans, 16 Ves. 512 (1809), sale of remainder after a life estate for an inadequate price; Earl of Portmore v. Taylor, 4 Sim. 182 (1831), assignment of annuities, subject to an outstanding life estate, for somewhat less than one-half their capitalized value.

[40] Berny v. Pitt, 2 Vern. 14 (1686); Wiseman v. Beake, 2 Vern. 121 (1690); James v. Oades, 2 Vern. 402 (1700).

[41] Berney's Case, 2 Swanst. 142 (1680); Earl of Ardglasse v. Muschamp, 1 Vern. 237 (1684); Bill v. Price, 1 Vern. 467 (1687); Lamplugh v. Smith, 2 Vern. 77 (1688); Freeman v. Bishop, 2 Atk. 39 (1740); Lamplugh v. Lamplugh, 1 Dick. 411 (1769).

power to bargain at the time of their greatest need.[42] Beyond this, the transactions themselves usually involved elements of risk which might justify a higher return than simple interest. In spite of these difficulties, doctrine rapidly hardened. The objective standard of fairness that equity enforced was a return of the money paid, plus simple interest. The factor of risk was very largely disregarded and emphasis was placed instead on the disproportionate gain resulting on the ancestor's timely death.[43]

In the nineteenth century conflict of opinion arose, not so much over the basic doctrine itself as over the method of determining the elements of a fair price. Whether the transaction was an outright sale or a mortgage of a reversionary interest, its fairness could not be measured merely by hindsight. The values exchanged in an outright sale or in the giving of security by a mortgagor would usually depend on uncertain future events, such as the death of a life tenant or a prospective failure of issue. It became increasingly common to express the tests of adequacy of consideration in terms of burden of proof, by throwing on the party resisting cancellation the burden of establishing the fairness of the whole transaction. But it was clear that this technique involved more than an initial procedural handicap.[44] If in the end the proof

[42] The point was strongly urged by the defendant in Twisleton v. Griffith, 1 P. Williams 310 (1716), but was answered by the court with the comment that sales of reversions "tended to the destruction of heirs sent to town for their education and to the utter ruin of families," and that inability to sell his reversion "might force an heir to go home and submit to his father or bite on the bridle and indure some hardships, and in the meantime he might grow wiser and be reclaimed."

In the nineteenth century several cases express serious misgivings as to the harmful effects on the prospective heir himself of too strict a regulation by equity. Shelly v. Nash, 3 Madd. 232 (1818); Bromley v. Smith, 26 Beav. 644 (1859).

[43] Berny v. Pitt, 2 Vern. 14 (1686), cancellation decreed in spite of a stipulation discharging the seller of a remainder interest of all obligation if he did not survive his father; Barnardiston v. Lingood, 2 Atk. 133 (1740); Crowe v. Ballard, 1 Ves. 214 (1790); Bowes v. Heaps, 3 Ves. & B. 117 (1814). An eddy in the main stream is Nichols v. Gould, 2 Ves. Sr. 422 (1752), where a "poor dragoon" was refused cancellation of a sale of a remainder after a fee tail, in spite of the death without issue only one month later of the tenant in tail. The court admitted that if there were "any degree of fraud or imposition" the sale would have been set aside, but concluded that the extreme uncertainty as to the value of the plaintiff's interest precluded a finding of inadequacy of price.

[44] Even as procedural handicap, this device had important effects in some situations. For example, in Salter v. Bradshaw, 26 Beav. 161 (1858), cancellation was decreed of a sale of a reversion made forty years before, on the ground that the purchaser had not preserved evidence that the price paid was at the time adequate. Again, in Benyon v. Fitch, 35 Beav. 570 (1865), the impossibility of proving the value of an interest which depended on the death of another person without issue male was held to require cancellation, since the purchaser had not sustained the burden of proof.

showed even a minor discrepancy between the values exchanged, cancellation was usually decreed.[45] Furthermore, though the use of a hindsight test was expressly disclaimed, actuarial calculations were used in some of the cases in place of a test of market value, measured by the price that other buyers would have been willing at the time to pay for the interest transferred.[46] This excess of zeal led to a reaction in the decisions[47] and finally to the intervention of Parliament, which provided in 1867 that "no purchase made *bona fide* and without Fraud or unfair Dealing of any Reversionary Interest in Real or Personal Estate shall hereafter be opened or set aside merely on the Ground of Undervalue."[48] But the effect of this legislation was negligible. It was construed, consistently with its language, to preclude cancellation *merely* on the ground of undervalue. Where elements of inequality between the parties were added, relief was freely given and even the burden of proof on the party resisting cancellation was preserved.[49]

Persistent efforts were made to reconcile these doctrines with main trends in other areas. The effort was only partly successful. The triumphant progress of laissez-faire was leaving its mark on nineteenth century equity, which increasingly relaxed its tests of adequacy of consideration in the interests of "freedom of contract."[50] But the protection given necessitous reversioners and remaindermen rested largely on inadequacy of price. In one aspect this special doctrine remained, as on its first appearance, an indulgence to a younger generation whose advancement in life was so long postponed by the family settlements of

[45] Newton v. Hunt, 5 Sim. 511 (1832); Edwards v. Burt, 2 De Gex, M., & G 55 (1852); Foster v. Roberts, 29 Beav. 467 (1861); Jones v. Ricketts, 31 Beav. 130 (1862); Nesbitt v. Berridge, 32 Beav. 282 (1863).

Even the cases in which the burden of proof was held to be sustained show the extremely close scrutiny to which these transactions were subjected and how narrow a margin might suffice for upsetting them. Headen v. Rosher, 1 McL. & Y. 89 (1824); Perfect v. Lane, 3 De Gex, F & J. 369 (1861).

[46] Particularly in Gowland v. De Faria, 17 Ves. 20 (1810), though the facts of that case showed a considerable discrepancy in values and the effect of Sir William Grant's language seems to have been exaggerated in later cases and by KALES, FUTURE INTERESTS, 2d ed., §§369 et seq. (1920).

[47] A test of market value, for what it was worth, was decisively established and the language of Gowland v. De Faria repudiated by the House of Lords in Earl Aldborough v. Trye, 7 Clarke & F. 436 (1840). Further citations are given by KALES, FUTURE INTERESTS, 2d ed., §§ 369 et seq. (1920).

[48] 31 & 32 Vict., c. 4.

[49] Tyler v. Yates, L.R. 6 Ch. App. 665 (1871); Earl of Aylesford v. Morris, L.R. 8 Ch. App. 484 (1871); In re Slater's Trusts, L.R. 11 Ch. Div. 227 (1879).

[50] E. g., Burrowes v. Lock, 10 Ves. 470 (1805); King v. Hamlet, 2 Mylne & K. 456 (1834).

the landed gentry.[51] The explanation given in the equity cases, however, was the combination of inadequacy of price with the economic necessity of the reversioner or remainderman. This emphasis on the element of economic necessity was justified by the economic pressures which were in fact revealed in most of the cases, though equity doctrines had tended to degenerate, in application, into mechanical formula.[52] Even with this emphasis, there remained the need for broader theory, which would relate this specialized form of relief to some larger objective of legal and equitable remedies. It was here that theory failed. About the best that could be suggested was that the cancellation of advantages secured through exploitation of need rested on "the unconscientious use of power."[53] As so often occurred before and since, the vagueness of the morality appealed to in equity cases obscured rather than clarified the functions of Chancery remedies.

This very vagueness of Chancery doctrines, however, helped greatly in their extension to related situations. The process of extension was

[51] The doctrine was primarily concerned, as in the earlier period, with outright sales of remainders and reversions in either real or personal property, numerous transactions of this type being included in the references collected in KALES, FUTURE INTERESTS, 2d ed., §§ 369 et seq. (1920). But it also included all attempts to impose charges, by way of mortgage or annuity, on reversionary or remainder interests. Gowland v. De Faria, 17 Ves. 20 (1810); Pennell v. Millar, 23 Beav. 172 (1856); Bromley v. Smith, 26 Beav. 644 (1859); Tyler v. Yates, L.R. 6 Ch. App. 665 (1871). Also surviving was the older doctrine as to *post obit* bonds. Wharton v. May, 5 Ves. 26 (1799, 1807); Marsack v. Reeves, 6 Madd. 108 (1821); Fox v. Wright, 6 Madd. 111 (1821); Bernal v. Marquis of Donegal, 3 Dow 133 (1814). The language of these cases indicated that the emphasis was not so much on the form of the transaction as on the fact that it was executed by a person "in the position of an expectant heir." Protection was in other words an incident to status.

[52] Most of the cases present personal histories of progressive financial disaster, with hounding by creditors and desperate efforts to postpone the collapse. Sometimes, however, the dealings with reversionary interests were perfectly straightforward business transactions, with the family solicitor and swarms of actuaries standing by. One illustration out of many is Bromley v. Smith, 26 Beav. 644 (1859), where the court pointed out at p. 664 that the transaction was not saved by the fact that the owner of a reversionary interest was a mature man of thirty-nine years who knew perfectly well what he was doing and added: "Neither is it necessary for the heir to shew that he was in distress for money at the time; that fact is assumed, from the circumstance of his having dealt with any on such a footing, and the assumption, that the person advancing the money has possibly taken advantage of that distress, is the reason why the Court throws upon him the burthen of proving that the bargain he executed was reasonable."

[53] Lord Selborne in Earl of Aylesford v. Morris, L.R. 8 Ch. App. 484 at 490 (1873), asserting that the decisions "raise 'from the circumstances or conditions of the parties contracting—weakness on one side, usury on the other, or extortion, or advantage taken of that weakness'—a presumption of fraud. Fraud does not here mean deceit or circumvention; it means an unconscientious use of the power arising out of these circumstances and conditions."

easiest in those cases where bargain transactions were made with persons who were physically and mentally weak. Such elements of personal disability occasionally appeared in the reversioner cases themselves, and in such cases the relations between doctrines as to sales of reversions and "undue influence" were particularly close.[54] The contribution of the reversioner cases is more directly traceable in other situations where disparity in worldly experience was coupled with extreme poverty or pressing economic need. The English cases, in considerable variety, which awarded cancellation, could spell out no reason more precise than "fraud," "surprise," or the "unconscionable" character of the bargain.[55]

[54] Bawtree v. Watson, 3 Mylne & K. 339 (1834).

Cases containing elements of "undue influence" but expressing a somewhat broader principle are Longmate v. Ledger, 2 Giff. 157 (1860), sale of fee simple interest in real estate for an inadequate price by a man seventy-two years old, formerly a farm laborer, "of weak and eccentric disposition" and without independent advice, these circumstances being said to throw on the purchaser the burden of proving that "no unfair advantage was taken of his weakness, and that a fair price was given to him" (id. at 163, 164); Clark v. Malpas, 4 De Gex, F. & J. 401 (1862), sale of fee simple interest by a man over sixty years of age, "in humble life . . . and unable to judge of himself the precautions to be taken in selling," and without independent advice (id. at 403, 404); Evans v. Llewellin, 1 Cox Ch. 333 (1787), sale of a half-interest in land by persons "in very mean circumstances, . . . and quite ignorant of their rights" (id at 335).

[55] Proof v. Hines, Cases Temp. Talbot 111 (1735), bond for £1000 given by a man illiterate and in desperate poverty in return for services in proving a claim to inheritance of an estate; Wood v. Abrey, 3 Madd. 417 (1818), tenant for life and remainderman join in sale of their interests so that the rule of the reversioner-remainderman cases does not apply, but inadequacy of price plus "the great distress" of the vendors held to justify cancellation; Pickett v. Loggon, 14 Ves. 215 (1807), sale of freehold and copyhold estates for an inadequate price by vendors in a state of "most abject and importunate distress"; Baker v. Monk, 4 De Gex., J. & S. 388 (1864), sale by a woman, aged sixty-seven, "in humble life, of slender education" to an experienced business man, at an inadequate price and without independent advice; Fry v. Lane, 40 Chan. Div. 312 (1888), sale of undivided interests in real and personal property for an inadequate price by "poor, ignorant men" without independent advice; James v. Kerr, 40 Chan. Div. 449 (1889), mortgage of an interest in real estate by a poor solicitor's clerk who was pressed by creditors and whose "position of poverty and necessity . . . put him practically at the mercy of the lenders so that the parties were not 'on equal terms.'" (P. 460).

In How v. Weldon, 2 Ves. Sr. 516 (1754), a sailor's assignment for an inadequate price of his share in prize money was set aside, sailors being described as "a race of men, loose and unthinking," so that plaintiff might be considered "at least in as favourable a light as a young heir" (id. at 518).

In Underhill v. Harwood, 10 Ves. 209 at 219 (1804), Lord Eldon asserted the jurisdiction of equity to review the adequacy of the consideration in a grant of an annuity in return for present cash, though the grant was made by business partners for the purpose of securing funds for use in their business. He summarized the effect of the equity cases by saying that "if the terms are so extremely inadequate as to satisfy the conscience of the Court by the amount of the inadequacy, that there must have been imposition, or that species of pressure upon distress, which in the view of this Court

Taken together they constituted an important departure from the class bias which first inspired the reversioner cases. They also established a source, from which American courts were to draw, for a broad doctrine that inadequacy of consideration would justify cancellation when coupled with extreme disparity in knowledge, experience, or economic and social position.

In the United States these doctrines as to sales of reversions underwent an important transformation. The main policy which had inspired the special treatment of reversioners and remaindermen, the policy of preventing dissipation of family estates, seemed wholly unacceptable to most American courts. Numerous cases rejected, sometimes with vigor, the English restrictions on sales of future interests and declared them as freely marketable as other forms of ownership.[56] Only in isolated instances was inadequacy of price admitted as a ground for attack on such transactions, and there the presence of other elements made it possible to explain results on broader grounds.[57]

On the other hand, the English doctrine was translated into a special control over sales of expectancies, redefined to include merely those bare expectations of inheritance which fell outside the common law scheme of vested and contingent estates. This result was in part produced by the continued refusal to admit that "ownership" could attach to such anticipations, so that attempts to transfer could be enforced only by way of specific performance of contract, after anticipations were realized through the death of the ancestor. The use of this technique made available the stricter tests of fairness and adequacy of price that

amounts to oppression, this Court would order the instruments to be delivered up, though Courts of Law might hold that judgment not within the sphere of their powers."

[56] Hayes v. Huddleson, 40 App. D.C. 183 (1913); McAdams v. Bailey, 169 Ind. 518, 82 N.E. 1057 (1907); Jaeschke v. Reinders, 2 Mo. App. 212 (1876); Davidson v. Little, 22 Pa. 245 (1853); Whelen v. Phillips, 151 Pa. 312, 25 A. 44 (1892); Cribbins v. Markwood, 13 Gratt. (Va.) 495 (1856); Provident Life and Trust Co. v. Fletcher, 237 F. 104 (1916).

Similarly, cancellation refused in transfers of legacies whose payment or enjoyment was postponed. Parmelee v. Cameron, 41 N.Y. 392 (1869); Jackson's Estate, 203 Pa. 33, 52 A. 125 (1902); Phillips' Estate, 205 Pa. 511, 55 A. 212 (1903); Singer's Estate, 217 Pa. 295, 66 A. 548 (1907).

[57] M'Kinney v. Pinckard's Exr., 2 Leigh (Va.) 167 (1830), extreme inadequacy of price plus economic necessity; Friedman v. Hirsch, 18 N.Y.S. 85 (1892), extreme inadequacy of price. The views of the English cases seem to be reflected in Dunn v. Chambers, 4 Barb. (N.Y.) 376 (1848), and Chambers v. Chambers, 139 Ind. 111, 38 N.E. 334 (1894), the dictrines of the latter case being considerably qualified by McAdams v. Bailey, cited in note 56.

are traditional in specific performance cases.[58] Another rationalization also appeared, the possible "fraud" on the ancestor if arrangements were made without his knowledge to alter the intended devolution of his estate. This suggestion, originating in Massachusetts, reflected some archaic conceptions of parental or ancestral authority that had been implicit in the English cases.[59] It led in a few states to a requirement of express consent by the ancestor to arrangements disposing in advance of any acquisitions from his estate.[60] This requirement was either rejected or ignored in most of the states.[61] Like the surviving requirement of adequacy of price, it must be explained chiefly through the special disadvantages which attach to the position of the "expectant" heir. American cases which maintain a strict control over sales of expectancies, as

[58] Numerous cases adopting this analysis are collected in 17 A.L.R. 597 (1922), 44 A.L.R. 1465 (1926), and 121 A.L.R. 450 (1939). In a few instances there has appeared an alternative mode of analysis, estoppel by deed operating on the title later acquired, when the attempt to transfer was by warranty deed. 17 A.L.R. 597 at 616 (1922), 44 A.L.R. 1465 at 1469 (1926), 121 A.L.R. 450 at 459 (1939). But see Mastin v. Marlow, 65 N.C. 696 (1871); Dailey v. Springfield, 144 Ga. 395, 87 S.E. 479 (1915).

Hostility to dealings with expectancies is carried to an extreme in Kentucky, where a series of cases have held that attempts to transfer such interests are absolutely void both at law and in equity, even when reinforced by covenants of warranty. 17 A.L.R. 615 (1922); 1 SIMES, FUTURE INTERESTS, § 234 (1936).

[59] In King v. Hamlet, 2 Mylne & K. 456 (1834), the parent's knowledge of his son's extensive dealings with his reversionary interests was held to eliminate the need for the protection ordinarily allowed. Earl of Chesterfield v. Janssen, 1 Atk. 301 (1750), mentions as one of several reasons for the rules protecting reversioners, the "fraud" on ancestors, in being "seduced to leave their fortunes to be divided among a set of dangerous persons, and common adventurers." Id. at 353.

The need for firm control over improvident youth seemed sufficiently strong to the court in Boynton v. Hubbard, 7 Mass. 112 (1810), so that a contract for sale of an expectancy was held void in an action at law for damages, on account of the "fraud" on the ancestor and the grim prospect that the heir might thereby be allowed to live in "prodigality, idleness, and vice."

[60] Farmers' Loan and Trust Co. v. Wood, 78 Ind. App. 147, 134 N.E. 899 (1922); Hight v. Carr, 185 Ind. 39, 112 N.E. 881 (1916); Stevens v. Stevens, 181 Mich. 438, 148 N.W. 225 (1914). In the two cases last cited, the requirement was enforced even though the ancestor was insane and therefore incapable of giving effective consent to the transaction.

[61] Gannon v. Graham, 211 Iowa 516, 231 N.W. 675 (1930); Hale v. Hollon, 90 Tex. 427, 39 S.W. 287 (1897); Fuller v. Parmenter, 72 Vt. 362, 47 A. 1079 (1900). Of the many cases that ignore this requirement, may be cited Clendening v. Wyatt, 54 Kan. 523, 38 P. 792 (1895); McDonald v. McDonald, 5 Jones (58 N.C.) 211 (1859); and Fidelity Union Trust Co. v. Reeves, 96 N.J. Eq. 490, 125 A. 582, 98 N.J. Eq. 412, 131 A. 208 (1924). And see Gadsby v. Gadsby, 275 Mass. 159, 175 N.E. 495 (1931).

that term is now redefined, express a continuing need for protective doctrine, though its scope is now narrowly limited.[62]

The broader ideas generated in the English reversioner cases encountered a different fate. Instead of being restricted in scope, as was the concept of the "expectancy," the tests of fairness in English equity cases were still further generalized. In the United States, as in England, the movement of ideas broadened out from the narrow channels within which it had been confined. The direction of this larger movement must now receive some attention, particularly with regard to modern judicial attitudes toward adequacy of consideration.

IV

LEGAL AND EQUITABLE DOCTRINES AS TO ADEQUACY OF CONSIDERATION

The Roman law rules of *laesio enormis*, allowing relief for inadequacy of price, have never achieved any general acceptance either at common law or in equity. Though undoubtedly known to some English students of Roman law, the rules of *laesio* were scarcely needed in the earlier centuries before the English law of contract had moved out from under the shadow of the penal bond. Nor is it a matter for regret that our law has not undertaken to solve the problem of the fair exchange by this purely arithmetical test.[63]

When assumpsit was developed, during the fifteenth and sixteenth centuries, as a remedy for breach of informal contract, the primary problem was to define in a technical formula the classes of transactions to which a technical system of remedies was being extended. The test of consideration, which began to take shape for this purpose in the sixteenth century, was not intended as a device for attack on unequal bar-

[62] The following cases are particularly suggestive as to the requirement of adequacy of price in transfers of expectant interests: Bridge v. Kedon, 163 Cal. 493, 126 P. 149 (1912); Richey v. Richey, 189 Iowa 1300, 179 N.W. 830 (1920); Klingensmith v. Klingensmith, 193 Iowa 350, 185 N.W. 75 (1922); Baltimore Humane Impartial Society v. Pierce, 100 Md. 520, 60 A. 277 (1905); Bayler v. Commonwealth, 40 Pa. 37 (1861); Norris' Estate, 329 Pa. 483, 198 A. 142 (1938); Butler v. Haskell, 4 Desaus. (S.C.) 651 (1816); Taylor v. Swafford, 122 Tenn. 303, 123 S.W. 350 (1909).

[63] The rules of *laesio* as they appeared in the Corpus Juris of Justinian allowed rescission to sellers of land where the price received was less than half the value of the property sold. Greatly expanded in the middle ages, the doctrine retained the purely arithmetical test of a 50 per cent discrepancy in value. The later experience of France and Germany with the doctrine is briefly sketched by Dawson, "Economic Duress and the Fair Exchange," 11 TULANE L. REV. 345 at 364 ff. (1937).

gains. Though the record is most incomplete, we can be sure that the alleviation of hardship and the overhauling of "hard" bargains were the functions of Chancery remedies. In the common law system the struggle over consideration was a struggle to express the results of procedural development in areas where policy and fairness required enforcement, not cancellation or revision. The early common law cases which refuse to review the "adequacy" of consideration must be read against this background.[64]

The refusal to review the "adequacy" of consideration came to rest on broader premises in the eighteenth and nineteenth centuries. As theories of economic individualism made their way into judicial thinking, interference with processes of price determination began to be described as not only impracticable but positively dangerous. Like any other form of governmental intervention, judicial attempts to fix a standard of equivalence seemed to hamper the enlightened pursuit of self-interest, which was expected to produce unlimited gains for society at large. The official doctrine of the common law of contract, reaffirmed with mounting conviction through the nineteenth century, excluded all inquiry into "adequacy," provided the transaction complied with some one of the various technical tests summed up in the term consideration. Even in the twentieth century, the fundamental postulates of economic individualism were nowhere asserted with such vigor as in judicial opinions in the field of private contract law. Considerations not only of practicability but of positive social policy seemed to exclude those ethical values which medieval thought had summed up in the concept of the "just price."

Even in the nineteenth century this broad conclusion was never completely accepted in fact. A standard of equivalence was enforced, for example, in equity cases involving specific performance. The explanation commonly given for this survival was the discretionary character of equitable remedies, reinforced by the suggestion that the refusal of

[64] Sturlyn v. Albany, Cro. Eliz. 67 (1587), is usually cited as early authority for the proposition that the adequacy of consideration would not be inquired into in common law actions. But in that case defendant, assignee of a lessee, promised plaintiff, the lessor, to pay the amount of overdue rent if plaintiff would show him a deed proving that the rent was due. Plaintiff accordingly did so, and then sued for the amount of the rent that was admittedly due, though defendant's own personal liability would have been hard to establish in the absence of the promise sued on. In this context large implications can scarcely be read into the court's statement that "when a thing is to be done by the Plaintiff, be it never so small, this is a sufficient consideration to ground an Action." The same comment may be made as to Bunniworth v. Gibbs, Style 419 (1654), which is likewise cited for the remark: "a little consideration will serve to ground a promise upon."

specific performance would have no serious consequences since a damage remedy was still available. This is not the place to explore the defects of the technique employed, defects that became more serious with the crystallization of the rules of contract damages in the nineteenth century.[65] It is enough to suggest that an ideal which was considered unsafe for wider use was here strongly asserted and brought to partial realization.[66]

In particular types of legal transactions a standard of equivalence was more vigorously maintained. A familiar example is the bargain between mortgagor and mortgagee for sale of the equity of redemption. The official reasons for strict control at this point are familiar. The economic necessities of the mortgagor, particularly after default in payment, leave him exposed to exploitation. Without denying him power to bargain his way out of financial embarrassment, courts of equity surround him with elaborate safeguards to ensure the fairness of the ultimate result.[67]

Closely connected with the rules as to sales of the equity of redemption are the rules evolved, particularly in equity, in the broader field of usury. It is from statute that we have derived the rules invalidating stipulations for excessive interest, as well as the particular standard of equivalence that defines a reasonable rate. But the experience of equity courts in awarding cancellation of usurious transactions familiarized them with an area in which an unequal exchange of values was frequently produced by the borrower's financial necessity. In early American cases there appeared some tendency to generalize the results of the English cases, and to relieve the distressed borrower under broader doctrines of case law.[68] The modern modifications of the usury

[65] The effect of refusing specific performance while a damage remedy is left available is discussed in 32 MICH. L. REV. 518 (1934).

[66] Here, as in other fields, many cases can be found which declare that inadequacy alone is no ground for refusal to enforce a contract, unless it "shocks the conscience" or other inequitable elements are present. Nevertheless it is clear that far stricter requirements are everywhere enforced in actions for specific performance and that many cases will refuse the remedy merely for inadequacy of consideration. Cases are collected in 65 A.L.R. 7 at 80 ff. (1930).

[67] 36 MICH. L. REV. 111 (1937), and cases collected in 129 A.L.R. 1435 at 1518 ff. (1940).

[68] Hough v. Hunt, 2 Ohio 495 (1826), contract to purchase land at a price more than twice its value, the purchaser being known to the vendor to be extremely pressed for money and the exorbitant price being attributed by the court to an "unfair advantage" taken of the purchaser's necessity.

Esham v. Lamar, 49 Ky. 43 (1849), mortgage of a slave for a period of eight years

statutes have restricted the growth of these remedial principles.[68] Nevertheless the usury statutes that have survived represent an important departure from the main policies of a competitive economy and are hard to reconcile, in detailed application, with the rules of a "free" market. The perennial problem of the small loan, with which modern legislatures are so much concerned, has pressed for solution because it combines the elements of urgent economic necessity and resulting disproportion in values exchanged.

In a survey of main attitudes toward the problem of equivalence, one should not neglect the special rules developed in equity for the so-called "confidential" and "fiduciary" relationships. Throughout this large and important group it is clear that ordinary processes of bargaining are considered inappropriate, because of unusual reliance on personal honesty and good faith that exists in fact or that is thought necessary for performance of "fiduciary" functions. In enforcing such expectations of honesty and good faith, the technique commonly used is the one already encountered in the English reversioner cases, a shift in the burden of proof. The person securing an advantage from a confidential or fiduciary relationship is required to prove not only his own full disclosure of all the relevant facts, but the fairness of the transaction as a whole. The attempt to frame substantive doctrines in terms of purely procedural handicap, together with the murky language employed in the equity cases, has disguised the main policies that are at work. Though discussion is largely in terms of false motivation ("fraud" and associated ideas), the effect is a standard of equivalence,

by settlers who had just removed to Kentucky and were destitute, the amount of the loan being equalled by the value of the slave's services in a single year and the contract being found to be "hard, unconscientious, and extorsive."

Deaderick v. Watkins, 8 Humph. (Tenn.) 520 (1847), assignment for less than half its value of part of a legacy due from an estate in process of distribution, the assignor being twenty-three years old and in urgent need of money.

A later case of the same type is Roux v. Rothschild, 37 Misc. 435, 75 N.Y.S. 763 (1902), though the result may be attributable to the usury statute. See also Bidwell v. Whitney, 4 Minn. (Gil. 45) 76 (1860); Sime v. Norris, 8 Phila. (Pa.) 84 (1871); In re Chicago Reed & Furniture Co., (C.C.A. 7th, 1925) 7 F. (2d) 885.

[69] Shriver v. Druid Realty Co., 149 Md. 385, 131 A. 815 (1926), refusing to modify a contract for the loan of money at excessive rates, where the usury statute in force expressly excepted corporate borrowers; Jenkins v. Moyse, 254 N.Y. 319, 172 N.E. 521 (1930). Other cases have refused, in a variety of situations, to find that the mere refusal to loan money was coercive within the tests for economic duress. Holmes v. Clark, 274 Ky. 349, 118 S.W. (2d) 758 (1938); McFarland v. Liberty Nat. Bank, 166 N.Y.S. 393 (1917); Starks v. Field, 198 Wash. 593, 89 P. (2d) 513 (1939). This issue will be further discussed in a later article.

that applies in a wide variety of personal relationships and to many types of legal transactions.[70]

In addition there survived some broader and still vaguer doctrines of equity which admitted mere inadequacy of consideration as ground for cancellation. The language used was characteristic. After denying that equity possesses its own independent tests of adequacy of price, it was usual to add that the result was otherwise where the inadequacy was so great as to "shock the conscience." In such cases the inadequacy was said to raise a presumption of "fraud" (often expressed as a conclusive presumption), so that relief could be rested on an ancient and respectable ground of equity jurisdiction, rather than on the doubtful ground of inadequacy as such.[71] Throughout the nineteenth century and down to the present time these ideas have persisted. They might be disregarded if there were not a considerable number of reported decisions in which such language was given effect through the grant of cancellation.[72] If criticism is to be levelled at such decisions, it should be

[70] For example, Kirby v. Arnold, 191 Ala. 263, 68 S. 17 (1915); Pye v. Pye, 133 Ga. 246, 65 S.E. 424 (1909); Reed v. Peterson, 91 Ill. 288 (1878); Walker v. Shepard, 210 Ill. 100, 71 N.E. 422 (1904); Swiney v. Womack, 343 Ill. 278, 175 N.E. 419 (1931); M'Cormick v. Malin,.5 Blackf. (Ind.) 509 (1841); Hunt v. Bass, 2 Dev. (17 N.C.) 292 (1832); Hutson v. McConnell, 139 Okla. 240, 281 P. 760 (1928); King v. Cohorn, 6 Yerg. (Tenn.) 75 (1834).

[71] The phrasing is that of Lord Eldon in Coles v. Trecothick, 9 Ves. 234 (1804), a specific performance case. Lord Eldon stated at p. 246 that specific performance should be refused where the inadequacy of price was "such as shocks the conscience, and amounts in itself to conclusive and decisive evidence of fraud. . . ." This phrasing was popularized by Story, as the test for determining the enforceability of contracts in equity. STORY, EQUITY JURISPRUDENCE, 3d ed., §§244-246 (1843). Most of the cases cited in notes 69, 71 and 72 use this language, with or without variations. In many others it appears as dictum.

[72] Kirby v. Arnold, 191 Ala. 263, 68 S. 17 (1915); Butler v. Duncan, 47 Mich. 94 (1881); Johnson v. Woodworth, 134 App. Div. 715, 119 N.Y.S. 146 (1909); Sherman v. Glick, 71 Ore. 451, 142 P. 606 (1914); Stephens v. Ozbourne, 107 Tenn. 572, 64 S.W. 902 (1901); Burch v. Smith, 15 Tex. 219 (1855). See also Mann v. Russey, 101 Tenn. 596, 49 S.W. 835 (1898).

A very large number of decisions have indicated their approval of the general language popularized by Story, using the "shock-the-conscience" and similar tests. These decisions are unsatisfactory as authorities in the sense that in practically all of them some elements of oppression, personal inequality or misleading conduct, short of positive misrepresentation, are found to be present in widely varying degrees. The following are merely illustrations: Lester v. Mahan, 25 Ala. 445 (1854); Shaffer v. Security Trust and Sav. Bank, 4 Cal. App. (2d) 707, 41 P. (2d) 948 (1935); Hardy v. Dyas, 203 Ill. 211, 67 N.E. 852 (1903); Marshall v. Billingsley, 7 Ind. 250 (1855); Matthis v. O'Brien, 137 Ky. 651, 126 S.W. 156 (1910); Barker v. Wiseman, 51 Okla. 645, 151 P. 1047 (1915); Birdsong v. Birdsong, 39 Tenn. 289 (1859); Kuelkamp v. Hidding, 31 Wis. 503 (1872); Carpenter v. Mason, 181 Wis. 114, 193

neither for the results reached (in granting cancellation with or without terms) nor for their failure to define the precise point at which the inequality of the exchange becomes sufficiently extreme. The ground for criticism should be the indirection which makes inadequacy of price seem a subsidiary element in judicial analysis. Confused as it is, the language of these cases contains a large element of truth, which may appear more clearly through a reformulation: an extreme disproportion in values in a bargain transaction requires explanation, and the explanation can usually be found in some misplaced reliance on the opposite party's good faith, some misleading partial disclosure, or some extreme inequality of the parties in knowledge, experience, or economic resources. If inequality in values is thus traced to its source in the conditions or the relations of the parties, the grant of judicial remedies seems no longer to endanger the economic foundations of an individualistic society. On the contrary, the function of judicial remedies becomes a policing function, the detection and correction of those factors which disturb and disrupt the "market." This function modern courts have shown themselves entirely willing to assume.[78]

The conclusion suggested by this short survey is that standards of equivalence have survived the hostile climate of the nineteenth century, though in disguised form. The language of the cases reflects the extreme reluctance with which a review of the "adequacy" of considerations is undertaken and in hundreds of cases such review has been refused. There remains a body of doctrine, however, to which courts are

N.W. 973 (1923). Numerous cases are collected in BLACK, RECISSION AND CANCELLATION, 2d ed., §§ 169-175, esp. § 175; and in L.R.A. 1916 D, 382.

A suggestive treatment of the whole subject, along lines similar to those above adopted, appears in 35 COL. L. REV. 1090 (1935).

[78] A review of appellate court decisions in the last five years reveals, if anything, an increased acceptance of the notion that inadequacy of consideration is itself a "badge of fraud." It is true of the more recent decisions, as of the earlier, that extreme inadequacy is almost always coupled with some other factor such as inequality in knowledge or education, economic necessity, reliance on personal good faith, or impaired capacity. It is plain, however, both from the language and the results of the cases that such additional factors, though in themselves insufficient, will be "seized hold of" when coupled with a striking disparity in values, and that the disparity in values is the primary basis for relief. Thomas v. Davis, 241 Ala. 271, 2 S. (2d) 616 (1941); Barner v. Handy, 207 Ark. 833, 183 S.W. (2d) 49 (1944); State Finance Co. v. Smith, 44 Cal. App. (2d) 688, 112 P. (2d) 901 (1941); Logue v. Von Almen, 379 Ill. 208, 40 N.E. (2d) 73 (1941); Elsasser v. Miller, 383 Ill. 243, 49 N.E. (2d) 21 (1943); Clement v. Smith, 293 Mich. 393, 292 N.W. 343 (1940); Tindel v. Williams, 187 Okla. 482, 103 P. (2d) 551 (1940); Downing v. State, 9 Wash. (2d) 685, 115 P. (2d) 972 (1941). Among the many cases in which similar language appears are Ross v. Koenig, 129 Conn. 403, 28 A. (2d) 875 (1942); Burroughs v. Mefford, 387 Ill. 461, 56 N.E. (2d) 845 (1944).

free at any time to appeal, allowing cancellation for an extreme inadequacy that "shocks the conscience" or constitutes "fraud." Even in the application of these vague standards, it is usually found that the disproportion in values is accompanied and explained by some particular circumstances of personal inequality or oppression, which the courts quite properly emphasize. Nevertheless, when the results already achieved in our law are added together, especially when the equity cases are taken into account as they usually are not, we find that the "adequacy" of consideration can be and frequently is measured by the techniques available in private litigation. It is no answer to say that courts can only reach the grosser instances of extreme disproportion in values or that disproportion in the values exchanged is usually produced by personal disparities between the contracting parties. That an ethical standard is limited or conditioned in its application is no proof that it has no place as a guide to judgment.

For a study of economic duress it is not necessary to explore the innumerable situations in which inequality in the values exchanged has inspired judicial attack on unfair bargains. For present purposes it may suffice to suggest that there is no single problem of equivalence, and that the problem of the unequal exchange must be approached in terms of the specific conditions which affect the bargaining power and the motivations of individuals in particular transactions. This mode of analysis requires a survey in each case, not only of the gains and losses involved in the transaction itself, but of the means employed in the bargaining processes that precede it.

It is at this point that attention should be directed again to the field of duress, which has been so much concerned with the problem of means. The first step is to suggest the specific connection between duress and the problem of the fair exchange.

V

ECONOMIC DURESS AND THE PROBLEM OF EQUIVALENCE

The historical connection between duress and the law of crime and tort has obscured the main function of duress doctrines, the prevention of unjust enrichment. This function existed from earliest times, for transfers extorted by the cruder forms of physical violence present the most obvious form of uncompensated and unjustified gain. But the absence of a developed theory of restitution and the limited scope of the remedies available postponed definition of the functions and the essential limitations of duress doctrines.

The broad limitation that emerges from the modern cases is that relief for duress will be given only for the purpose of restoring the excess over what is reasonably and justly due, and to the extent that such excess is shown to exist. This limitation, fundamental as it is, is ordinarily taken for granted. In the great mass of duress cases, restitution is sought of transfers or payments made in settlement of a non-existent debt. Retention of the money or property transferred is assumed to involve an uncompensated enrichment of the transferee; the debate is concerned merely with the means by which the transfer was coerced. In a relatively small percentage of the cases, however, the issue is directly presented, through a claim by the transferee that his retention of the money or property transferred is justified because some or all of the debt in question was owed in fact. Here the cases are unanimous in holding that recovery must be limited to the amount of the over-payment, even though the means by which the payment was induced are assumed to be improper. Most frequently the question is raised in actions of quasi-contract, and the result can then be explained in terms of the "equitable" bases of the quasi-contract remedy.[74] But the same result is also reached in other forms of action, at law and in equity, though the reasons given are sometimes quite obscure.[75]

[74] An example is Beckwith v. Frisbie, 32 Vt. 559 (1859), an action of assumpsit for money paid to release a cargo of oats detained by defendants. Duress of goods was found, but recovery was reduced by the amount which represented the reasonable expense of storing the oats through the winter, an expense which the plaintiff as owner was required by "equity" to assume. Similarly, McVane v. Williams, 50 Conn. 548 (1883); City of Chicago v. Malkan, 119 Ill. App. 542 (1905); Koenig v. Peoples Gas Light and Coke Co., 153 Ill. App. 432 (1910); Foss v. Whitehouse, 94 Me. 491, 48 A. 109 (1901); Congdon v. Preston, 49 Mich. 204, 13 N.W. 516 (1882); Minor Lumber Co. v. City of Alpena, 97 Mich. 499, 56 N.W. 926 (1893); Briggs v. Boyd, 56 N.Y. 289 (1874); Richmond v. Union Steamboat Co., 87 N.Y. 240 (1881); C. W. Hahl and Co. v. Hutcheson, (Tex. Civ. App. 1917) 196 S.W. 262; Dale v. Simon, (Tex. Civ. App. 1923) 248 S.W. 703; Guetzkow Brothers Co. v. Breese, 96 Wis. 591, 72 N.W. 45 (1897); The John Francis, (D.C. Ala. 1911) 184 F. 746; Champ Spring Co. v. United States, (C.C.A. 8th, 1931) 47 F. (2d) 1.

The numerous cases where a similar limitation on recovery was assumed without discussion, may be illustrated by Cunningham v. Hussey, 122 Me. 565, 119 A. 869 (1923); Jones v. Sherwood Distilling Co., 150 Md. 24, 132 A. 278 (1926); Link v. Aiple-Hemmelmann Real Estate Co., 182 Mo. App. 531, 165 S.W. 832 (1914).

See also Tisdale v. Bryant, 38 Cal. App. 750, 177 P. 510 (1918); Young v. Hoagland, 212 Cal. 426, 298 P. 996 (1931); Cazenove v. Cutler, 4 Metc. (45 Mass.) 246 (1842); Nutting v. McCutcheon, 5 Minn. 310 (1860); Powell v. Grand Lodge, 349 Mo. 955, 163 S.W. (2d) 1038 (1942); Upshaw v. Mutual Loan Assn., 29 Misc. 143, 60 N.Y.S. 242 (1899); Mee v. Town of Montclair, 84 N.J.L. 400, 86 A. 261 (1913); Coleman v. Merchants' Nat. Bank, 6 Ohio Dec. Reprints 1063 (1881).

[75] Diller v. Johnson, 37 Tex. 47 (1873), defense of duress not admissible in action on promise to pay plaintiff's debt to a third party, since there was no showing

Not all the duress cases involve compulsion to pay a supposed debt. Frequently the transaction induced by duress will contemplate an exchange of values through payments made, property transferred, or services rendered by the coercing party. Here the inequality of the exchange may be less striking and extreme. However, the machinery of rescission and restitution will produce a restriction on relief for duress that is strictly comparable to the restriction above suggested. As a condition to rescission the party under duress will be required to tender or account for the value received by him.[76] Or if judicial revision of the contract is attempted, restitution remedies will usually aim only at the excess value that was extorted through duress.[77] In either event the net result is to confine restitution to the margin of gain received by the coercing party and thus redress the inequality in the exchange

It is not meant to suggest that inequality in the values exchanged is an indispensable element in relief for duress. Where, for example, the transaction itself did not involve economic elements, as in a contract of marriage, neither the tests for duress nor the remedies employed can make use of economic tests.[78] Furthermore, where extreme forms of

that defendant was not independently liable for the amount in question; Doolittle v. McCullough, 7 Ohio St. 299 (1857), recovery in trover for the value of property transferred by plaintiff under threats of physical violence reduced by the amount of plaintiff's debts to third persons, paid off with proceeds of property transferred; Tisdale v. Bryant, 38 Cal. App. 750, 177 P. 510 (1918), refusal to discharge lien on real estate held not improper pressure since object was to compel payment of debt actually owed, though previously released. Connecticut Mutual Life Ins. Co. v. Stewart, 95 Ind. 588 (1884), is also interesting.

[76] Royal v. Goss, 154 Ala. 117, 45 S. 231 (1907); McCoy v. James T. McMahon Construction Co., (Mo. 1919) 216 S.W. 770; Graham v. Fisher, 244 App. Div. 740, 278 N.Y.S. 982 (1935); Carroll v. Fetty, 121 W. Va. 215, 2 S.E. (2d) 521 (1939). Where payments have been received for a release of a cause of action, tender of restitution may of course be dispensed with in some instances and the amount received merely deducted from the recovery on the released cause of action. Perkins Oil Co. v. Fitzgerald, 197 Ark. 14, 121 S.W. (2d) 877 (1938); Hazelhurst Oil Mill & Fertilizer Co. v. United States, (Ct. of Claims 1930) 42 F. (2d) 331 (1930).

[77] Lanigan v. Scharton, 238 Mass. 468, 131 N.E. 223 (1921); Coveney v. Pattullo, 130 Mich. 275, 89 N.W. 968 (1902); Shafer v. Giese, 135 Wash. 464, 238 P. 3 (1925); Barnett Oil & Gas Co. v. New Martinsville Oil Co., 254 F. 481 (1918), duress being supplemented here by misrepresentation as ground for judicial revision of the contract. There are few actions reported like that in Dennehy v. McNulta, (C.C.A. 7th, 1898) 86 F. 825, in which price revision was sought by the buyer in a sale of goods. Relief for duress was there refused, but it is significant that the only relief requested was recovery of the excess over a reasonable price, the excess being extracted as plaintiff claimed through defendant's monopoly of the supply.

[78] As in Quealy v. Waldron, 126 La. 258, 52 S. 479 (1910), and Fowler v. Fowler, 131 La. 1088, 60 S. 694 (1913). But even here, when means less drastic than

coercion have been used, a broad policy may dictate the grant of rescission remedies without regard to the inequality of the exchange.[79] Such cases, however, are relatively rare. Furthermore, if the transaction in question is equally advantageous to both parties, there will seldom be a motive either for exerting pressure to induce it or for securing its cancellation when the pressure is removed. The fact situations toward which duress doctrines are directed are, overwhelmingly, situations in which an unequal exchange of values has been coerced by taking advantage of a superior bargaining position. The thesis now advanced is that the restitution remedies employed in modern law are aimed merely at cancelling out the gain, by direct or indirect means.[80]

The question here considered has been most acute in cases of duress through threats of criminal prosecution. In no other field of duress has doctrine been so confused through conflict between the desire to ensure a reasonable degree of freedom for individual choice and the desire to enforce fair settlements of existing liability. The older views that made threats of prosecution available as a means of enforcing civil satisfaction have now been largely superseded. The extreme disadvantages of a person accused of crime, and threatened with arrest and public exposure, have led the modern cases to a wide extension of remedial principles. This development has been fitted into the classic conception of duress by describing the pressure as inherently improper, since proc-

father's shotgun have been employed (e.g., threat of criminal prosecution), an enforced marriage may sometimes seem a justified form of specific reparation. Rogers v. Rogers, 151 Miss. 644, 118 S. 619 (1928).

[79] For example, in Sun-Maid Raisin Growers v. Papazian, 74 Cal. App. 231, 240 P. 47 (1925), no showing was required that the price to be paid for defendant's raisin crop was unduly low, where it appeared that the agreement to sell the crop was induced by an organized program of physical violence, with night-riders and other trimmings. Even with less extreme forms of coercion, the policy of protecting individual freedom of choice may sometimes provide in itself a sufficient ground; an illustration is the high pressure salesmanship used on a nervous and vulnerable woman in Schoellhamer v. Rometsch, 26 Ore. 394, 38 P. 344 (1894), though here as usual the unfairness of the bargain was a principal ground for the relief given.

[80] This contention is of course not wholly applicable to those very exceptional cases where the relief is given for duress not by way of restitution but essentially on a theory of damages for tort. Illustrations are Slade v. Slade, 310 Ill. App. 77, 33 N.E. (2d) 951 (1941); Smith v. Blakesburg Savings Bank, 182 Iowa 1190, 164 N.W. 762 (1917); Neibuhr v. Gage, 99 Minn. 149, 108 N.W. 884 (1906); News Publishing Co. v. Associated Press, 114 Ill. App. 241 (1904); White v. Scarritt, 341 Mo. 1004, 111 S.W. (2d) 18 (1937); and perhaps Carew v. Rutherford, 106 Mass. 1 (1870). Compare Bancroft v. Bancroft, 110 Cal. 374, 42 P. 896 (1895), refusing to entertain an action for damages for undue influence. The subject is discussed in 39 HARV. L. REV. 108 (1925).

esses intended for the vindication of public interests are misapplied when used to enforce a civil liability.[81] It may be questioned whether the ideal so vigorously expressed conforms to the ordinary practice either of law enforcement officials or of private individuals who are damaged by criminal acts.[82] The difficulty in maintaining it is most apparent where the imminence of criminal prosecution (whether or not "threats" were used) has led to a settlement of a proved or admitted civil liability. Here numerous courts have refused to intervene, offering reasons which are hard to reconcile with their views in other situations.[83] Even more relevant for present purposes are the cases where some civil liability is proved but the threat of prosecution has been used to extract somewhat more than was actually due. Here several courts have specifically held that remedies for duress, both at law and in equity, must concern themselves only with the excess over the proved liability.[84] Where the pressure used is severe and the amount of the civil liability is still in doubt, there is good reason for allowing recission without attempting to determine precisely the civil consequences of the alleged crime. But with this qualification it can be asserted that in threats of criminal prosecution, as in other types of duress, the fairness of the settlement is a crucial question, both in determining whether rescission should be granted and in defining the scope and purpose of the remedies granted.[85]

[81] Morse v. Woodworth, 155 Mass. 233, 27 N.E. 1010 (1891), and cases cited by 5 WILLISTON, CONTRACTS, rev. ed., § 1611 (1937).

[82] See the instructive comment, "Restitution and the Criminal Law," 39 COL. L. REV. 1185 (1939).

[83] 5 WILLISTON, CONTRACTS, rev. ed., §§ 1612 and 1615 (1937).

[84] Wilbur v. Blanchard, 22 Idaho 517, 126 P. 1069 (1912); Heaps v. Dunham, 95 Ill. 583 (1880); Rood v. Winslow, Walker Ch. (Mich.) 340 (1844); Briggs v. Withey, 24 Mich. 136 (1871). See also Smith v. Thomas, 185 Ark. 613, 48 S.W. (2d) 561 (1932); Kohler v. Wells, Fargo, & Co., 26 Cal. 606 (1864); Kronmeyer v. Buck, 258 Ill. 586, 101 N.E. 935 (1913); Beath v. Chapoton, 115 Mich. 506, 73 N.W. 806 (1898); Clement v. Buckley Mercantile Co., 172 Mich. 243, 137 N.W. 657 (1912); Allen v. Leflore County, 78 Miss. 671, 29 S. 161 (1900); Fry v. Piersol, 166 Mo. 429, 66 S.W. 171 (1901); Rostad v. Thorsen, 83 Ore. 489, 163 P. 987 (1917).

[85] The carefully guarded statement in 5 WILLISTON, CONTRACTS, rev. ed., § 1616 (1937), is significant. After an elaborate and effective argument in favor of extension of relief against threats of criminal prosecution, the conclusion is that "In spite of the arguments advanced in the preceding section, it seems better to hold that *where anything other than a satisfaction of the precise civil obligation is obtained* by coercion through threats of prosecution by the creditor, the transaction should be avoided without reference to its reasonableness, and excellent authority supports this view." (Italics supplied.)

The strongest statement in the cases in support of this position is in Fidelity &

The analysis just suggested is elementary, so elementary that it is usually overlooked. Nevertheless, the specific connection between duress and the problem of the fair exchange not only helps to clarify particular situations but it has much larger implications. If it can be assumed that the object of relief for duress is to cancel out advantages secured by superior bargaining power, the whole group of duress cases takes on a new perspective. The objective of ensuring the freedom of the individual will, so frequently proposed in the nineteenth century cases, becomes on this analysis an incidental or at most a subsidiary objective. More important, the concentration of the modern cases on distinction between legal and illegal means seems misdirected, a survival from an earlier period in which duress doctrines were merely an adjunct of the law of crime and tort.

It is indeed this concentration on distinctions between legal and illegal means which has chiefly arrested the modern development of the law of duress. No single formula has achieved so wide a circulation in the duress cases as the statement that "It is not duress to threaten to do what there is a legal right to do." Certainly no other formula is anything like so misleading. Its vice lies in the half-truth it contains. For an enormous range of conduct is included in the class of acts that there is a "right" to do (and therefore, under this formula, to threaten). At one extreme are various types of severe injury and oppression that narrowly escape the sanctions of the law of crime and tort. At the opposite extreme are the types of pressure that are specifically provided by organized legal agencies, for the very purposes for which they are used (e.g., the remedies of civil litigation). Somewhere between these extremes must be classed all those multiplied forms of economic pressure

Guaranty Co. v. Cook, 43 Wyo. 356, 5 P. (2d) 294 (1931). But this case illustrates the real difficulty. Plaintiff sued for restitution of money paid in settlement of an alleged bank shortage; the jury's verdict established that whatever loss had occurred was not the result of intentional misappropriation; defendant sought nevertheless to establish its claim that a loss had occurred through plaintiff's negligence. Not only was this latter claim unsupported by evidence of negligence, but the court indicated a strong reluctance to throw on plaintiff the burden of litigating this doubtful and difficult question, as a condition to restitution of payments which had been coerced by harsh and unfair means. The problem thus narrows to a problem in burden of proof; and a strong showing can properly be demanded of a creditor who seeks to justify retention of money or property extorted by extremely oppressive means. Particularly should this be so where the threat was to prosecute for a crime unrelated, or only remotely related, to the civil liability asserted. This last point is still better illustrated by Thompson v. Niggley, 53 Kan. 664. 35 P. 290 (1894), which hardly supports the proposition for which it is cited by Patterson, that "In most of the cases . . . the fairness of the settlement has been treated as immaterial." 2 PATTERSON, CASES ON CONTRACTS 64, note 4 (1935).

by which the exchange of goods and services is accomplished in an individualist society. Without doubt such forms of pressure are normally permissible, and it is the essence of economic individualism to subject their use to a minimum of external regulation. From this it by no means follows that the effects of pressure exerted in particular cases will always escape judicial scrutiny. Doctrines of duress are intended to raise precisely the question whether it is "rightful" to use particular types of pressure for the purpose of extracting an excessive or disproportionate return. Over the whole range of conduct to which this question applies, it is plain that the tests of the criminal law or a damage remedy can no longer determine the limits of relief for unjust enrichment. The insight of Holmes cut through to the central distinction: "When it comes to the collateral question of obtaining a contract by threats, it does not follow that, because you cannot be made to answer for the act, you may use the threat."[86]

VI

CONCLUSION

The preceding survey was intended to suggest that the modern American law of duress reflects the convergence of several lines of growth, originally moving from sources quite distinct. The symptom of this convergence has been an increasing interplay and transfer of ideas. Its result has certainly not been a coherent body of doctrine, unified around some central proposition; on the contrary, the conflict and confusion in results of decided cases seem greater than ever before. This conflict and confusion must be attributed in part to the fact that the

[86] Silsbee v. Webber, 171 Mass. 378, 50 N.E. 555 (1898). In the very useful articles by Dalzell, "Duress by Economic Pressure," 20 N.C. L. REV. 237, 341 (1942), the position ultimately adopted is substantially similar to the one suggested above. Unfortunately at the outset of his review of the decisions, the author proposes a definition of duress which restates conventional doctrines and includes the "wrongfulness" of the threat as one of the two essential elements (id. at 240). The same requirement is presupposed in much of the later discussion. However, in the concluding section it is pointed out (id. at 341, 361-7) that the test of "wrongfulness" should not be identical with the tests of a damage remedy, since the object of relief for duress is restitution, not damage liability or penalty. It is finally concluded that a threat should be held "wrongful" for purposes of duress doctrines if it involves either "(1) a threat to commit an actionable wrong; (2) a threat to misuse a legal power given for other legitimate ends; (3) a threat to maintain a lawsuit or defense which ultimately proves to be unsustainable; (4) a threat to violate the standards of decent conduct in the community." The whole discussion at this point and in the final passages (id. at 341, 384-6) is helpful in indicating the connection between duress doctrines and the problem of the fair exchange.

processes of growth are still continuing and the effects of earlier history are not yet dissipated. In part, however, they are due to the complex issues of ethics and economic policy that constantly intrude themselves and on which courts, like other agencies of organized society, must take a positive stand.

It is evident that courts have neither the equipment nor the materials for resolving the basic conflicts of modern society over the distribution of the social product and the limits to be set to the use, or misuse, of economic power. The issues involved in these conflicts must be reserved, for the most part, for decision by other means. The limited range of judicial action makes it no less important, however, that the issues raised in private litigation be correctly identified and placed in proper perspective. Above all, where judicial review is refused the decision should not depend on a formula inherited from the thirteenth century but on a conscious evaluation of the factors that make judicial review impracticable or unwise.

The history of generalization in this field offers no great encouragement for those who seek to summarize results in any single formula The direct conflict in decisions, on facts substantially identical, makes it likewise impossible to formulate any general proposition that could now achieve anything like universal acceptance. Nevertheless, it seems clear that many decisions have already shifted a considerable distance beyond the limits defined by conventional statements of doctrine and that further shifts are to be expected. The most that can be claimed is that change has been broadly toward acceptance of a general conclusion—that in the absence of specific countervailing factors of policy or administrative feasibility, restitution is required of any excessive gain that results, in a bargain transaction, from impaired bargaining power, whether the impairment consists of economic necessity, mental or physical disability, or a wide disparity in knowledge or experience.

A conclusion so broadly stated leaves open innumerable questions that need to be examined in detail. There remain, for example, the important questions so much discussed in the field of economic duress, as to the nature and propriety of the means of pressure used. There remain also some difficult problems in determining the existence of pressure in fact, especially in relation to alternative means that may exist to relieve it. Most important of all are the factors of policy or administrative feasibility which may lead to refusal of judicial relief, when other necessary elements are present. Some of these questions will be further discussed in a subsequent article.

The shift in emphasis that is now proposed involves the assumption

that our courts cannot remain indifferent, in fact are not indifferent, to excessive and unjustified gains that are directly traceable to disparity in bargaining power. This assumption does not involve any expectation that the methods of private litigation will be used to overhaul the immense range of transactions involving the sale or exchange of goods and services in a competitive society. The factors that lead to judicial abstention are themselves basic; but it is time that they be examined.

THE ENGLISH COMMON LAW CONCERNING MONOPOLIES

WILLIAM L. LETWIN†

IT HAS BEEN WIDELY BELIEVED that the common law always favored freedom of trade. When English and American judges during the eighteenth and nineteenth centuries decided cases against monopolists, engrossers, or restrainers of trade, they thought they were continuing a tradition that reached back into "time of which no man hath memory." The congressmen who drafted and passed the Sherman Antitrust Law thought they were merely declaring illegal offenses that the common law had always prohibited. Those judges and legislators, like other lawyers, must have known, or at least would not have doubted, that the common law rules on these subjects had changed in the course of time, for it is taken as axiomatic that the common law "grows." But it is not always recognized that the common law can change its direction, and without much warning begin to prohibit practices it had formerly endorsed, or to protect arrangements it had earlier condemned. Lawyers do not so readily see that the common law at any given time reflects the economic theories and policies then favored by the community, and may change as radically as those theories and policies. As a result they have too easily accepted the mistaken view that the attitude of the common law toward freedom of trade was essentially the same throughout its history.

But the common law did not always defend freedom of trade and abhor monopoly. For a long time it did quite the opposite: it supported an economic order in which the individual's getting and spending were closely controlled by kings, parliaments, and mayors, statutes and customs, and his opportunities limited by the exclusive powers of guilds, chartered companies, and patentees. The common law first began to oppose this system of regulation and privilege at the end of the sixteenth century; it did not do so wholeheartedly until the eighteenth century; and by the middle of the nineteenth century, it had again lost its enthusiasm for the task. It would have been surprising if the pattern of development had been different. Changes in the common law are changes in the attitudes of judges and of lawyers; it would have been remarkable if they had persistently opposed monopoly when the rest of the community did not know

† Research Associate, University of Chicago Law School.

the word and considered the phenomenon natural or desirable. It would have been strange if lawyers had upheld laissez-faire policies centuries before any statesman or economist had advocated or stated them, and had continued following them long after they had been abandoned or denied by the rest of the community. In fact, English laws governing monopoly and English policies for the economic organization of society changed together, except for minor differences in timing. The English law of monopoly traditionally includes four branches: the law on monopoly proper, whether by patent, charter, or custom; on forestalling, engrossing, and regrating; on contracts in restraint of trade; and on combinations in restraint of trade. These branches, distinct in form and based on more or less independent bodies of precedent, nevertheless show the same development from an active support of monopolies in the earliest period, through active opposition during an interlude of less than two centuries, to the leniency and indifference which characterized them in 1890.

I

The idea that the common law opposed monopolies from the earliest time onward was invented largely by Sir Edward Coke,[1] who argued that monopoly was forbidden by the Civil Law,[2] and implicitly by Magna Carta[3] as well as by certain statutes of Edward III's reign.[4] But the earliest common-law precedent he could mention was a case that arose during the fourteenth century, and the modern lawyers and historians who follow his authority continue to cite that case as evidence of the ancient antagonism of common law to monopolies.[5] Yet the case gives at least equally good evidence to the contrary. One John Pecche had a patent giving him the exclusive right to sell sweet wines at retail in London, and the Parliament of 1376 petitioned that he be punished for the flagrant use he had made of his privilege, "to the great damage and oppression of the people."[6] Had this been an action to suppress and punish monopoly, it would tend to justify the theory that Coke put forward. But the fact that Pecche, who had a monopoly, was punished, does not mean that he was punished for having the monopoly.

[1] Wagner, Coke and the Rise of Economic Liberalism, 6 Econ. Hist. Rev. 30 (1935).

[2] See Coke's argument in Davenant v. Hurdis, Moore* 576,* 580 (K.B., 1599).

[3] 2 Coke, Institutes 47, 62–63; 3 ibid., at c. 85.

[4] 3 ibid., at c. 85.

[5] 3 ibid., at 181; 11 Co. Rep. *53, *88a, b. Cf. 4 Holdsworth, History 344 n. 6 (1924). The case was cited in the same sense by Laurence Hyde during the parliamentary debate on monopolies in 1601. Tawney and Power, 2 Tudor Economic Documents 275 (1924).

[6] Rotuli P. rum, 50 Edw. III, No. 33 (1376).

Pecche came by his patent during the last years of Edward III's reign, when the King was too old to rule actively, and his heir, the Black Prince, was too ill. Another son, John of Gaunt, Duke of Lancaster, was therefore virtual ruler of England from about 1370 to 1377. He hoped to succeed his father, but until that happy time came he intended to make the best of his position, and so he set out to increase his fortune. His chief agent in this venture was Richard Lyons, a merchant and alderman of London, who apparently financed Gaunt, in return for which the Duke gave or sold economic privileges to him and to his associates.[7] John Pecche, merchant, alderman, and once mayor of London,[8] was among the associates favored. In 1373, "by the assent and aid of Richard Lyons,"[9] he was given letters patent permitting him and his deputies to sell sweet wines at retail in London, notwithstanding an ordinance of Parliament prohibiting retail sale of sweet wines throughout the realm.[10] In return for this exclusive privilege,[11] he was to pay the King a fee of 10 shillings on each pipe of wine he sold.[12] The King further endorsed the grant by sending writs to the Mayor and Sheriffs of London commanding them to give Pecche every assistance in exercising his monopoly, and ordering the Mayor to set a reasonable price for the wine.[13] Nor did the Mayor and Aldermen of London seem to object in principle to the monopoly; in fact Pecche asserted that they had approved it before he put it in practice.[14] Needless to say, they denied it in Parliament, but however they may have felt about Pecche's privilege, they were not generally averse to monopoly: only a few years earlier they had leased to Lyons the three taverns reserved for the sale of sweet wines, at an annual rent of £200,[15]—which suggests that they did not object to monopolies that paid them good returns.

[7] Trevelyan, England in the Age of Wycliffe 10–12 (1915 ed.). For details on Lyons, see Sargeant, The Wine Trade with Gascony, in Unwin (ed.), Finance and Trade under Edward III, 297–98 (1918).

[8] 1 Riley, Memorials of London 308; 2 ibid. 390 (1868).

[9] Rot. Parl., 50 Edw. III, No. 33 (1376).

[10] The patent is missing from the Calendar of Patent Rolls, but is given in the Calendar of Letter Books of the City of London, Letter Book G, 318 (Nov. 30, 1373). The prohibitory ordinance of Parliament seems to be missing both from Rot. Parl. and Statutes of the Realm, but is referred to in Rot. Parl. when repealed (cf. note 21 infra) as well as in the patent.

[11] Exclusive only in London; John Beverle had a similar license for Boston, Lincoln, Staunford, and Grantham. C.P.R., 48 Edw. III, 414 (March 5, 1374).

[12] Rot. Parl., 50 Edw. III, No. 33 (1376).

[13] Letter Book G, 318 (Dec. 13, 1373); ibid., at 320 (Dec. 11, 1373).

[14] Rot. Parl., 50 Edw. III, No. 33 (1376). Letter Book H, 38–40 (Aug. 1, 1376).

[15] Letter Book G, 199 (Aug. 26, 1365).

For three years Pecche worked his monopoly in peace. But John of Gaunt became less popular than ever, encountering increased opposition from those who preferred that the Black Prince's son, Richard, should succeed Edward, from the Church, because he sponsored Wycliffe, and from the Londoners, who hated his arrogance and showed their antagonism finally by burning his palace during the Rebellion of 1381. When Parliament met in 1376, it therefore attacked Gaunt, but indirectly, by accusing his associates. It attainted Lyons for engrossing and raising the prices of "all the merchandise that came into England," William Lord Latimer for taking bribes and misappropriating funds, Alice Perrers, the King's mistress, for obstructing justice,[16] and Pecche, for having fraudulently obtained and excessively exploited his patent, and for having failed to pay the King the required fees. Pecche was sentenced "to be imprisoned, to make fine and ransom to the King, and also to give satisfaction to the parties complaining of his extortionate prices."[17] He spent only a short while in prison,[18] was soon after given a pardon excusing him from all further penalties and from paying any outstanding license fees,[19] and although his patent was not revoked,[20] it lost its value when Parliament repealed the former prohibition so that thereafter anyone could sell sweet wine at retail.[21]

Pecche's case was an accident of contemporary politics rather than part of an already developed common-law tradition against monopolies. The ultimate object of the attack was Gaunt; his various agents were assailed on the most plausible pretexts available. Parliament could find no other form in which to accuse Pecche except that he had defrauded the King and charged the public unreasonable prices. The word monopoly was not even mentioned in the accusation; it was unknown in England until over a century later.[22] The legal concept then existing which came closest to the notion of monopoly was "engrossing"; Lyons was accused of engrossing,

[16] 2 Longman, Life and Times of Edward III 251–54 (1869).

[17] Rot. Parl., 50 Edw. III, No. 33 (1376).

[18] It cannot have been long, for Parliament met on April 24, 1376 and Pecche was released from the Tower on an order of mainprise dated July 26; Calendar of Close Rolls, 50 Edw. III, 437 (1376).

[19] C.P.R., 51 Edw. III, 448, 457 (April 10, 1377).

[20] As Wagner, op. cit. supra note 1, at 41, points out, following Gordon, Monopolies by Patents 231, n. (1897). Both, however, overlooked the condition that made revocation unnecessary.

[21] Rot. Parl., 50 Edw. III, No. 14 (1376). Cf. C.C.R., 51 Edw. III, 529 (Feb. 18, 1377).

[22] It was first used, according to Fox, Monopolies and Patents 24 (1947), by Thomas More in 1516.

but Pecche could not be accused of it because he had acted under royal license. Parliament did not question the King's right to grant such patents; indeed a judge during the same reign declared that the King might grant privileges "even though, prima facie, they appear absolutely against common right."[23] There was as yet no basis at common law for holding a grant void because it created a monopoly.

The great movement against the granting of monopolies by letter-patent began only at the end of the sixteenth century, although it was so strongly supported that within less than a hundred years the principle had been established that Parliament alone could grant a monopoly, and that generally even it could not, as the King had regularly done, sell a patent or award it on a whim or as a friendly gesture. By the end of the seventeenth century the royal letter-patent had been converted into a more or less modern version of the patent, justifiable only by a solid contribution to economic development. The process was not, however, moved by coherent opposition to monopoly; it was brought about mainly by disturbances within the monopolistic system administered largely by the guilds, and by objections not to the broad economic effect of monopolies but to the political power which the crown exercised in granting them.

The first recorded case on monopolies was *Davenant v. Hurdis*, or *The Merchant Tailors'* case decided in 1599,[24] which shows not only the extent of monopolistic control that the guilds exercised, but also the ends that such controls were supposed to serve, and the collisions that were taking place between several guilds, as each tried to maintain intact its power over a trade. The case arose under a by-law passed by the London tailors' guild in 1571, titled "An Ordinance for Nourishing and Relieving the Poor Members of the Merchant Tailors Company." The ordinance begins with a noble preamble, "Forasmuch as it is the duty of every Christian society to help and relieve every willing labouring brother in the Commonwealth, and especially such as are incorporated, grafted, and knit together in brotherly society . . . ,"[25] and goes on to require every merchant who belongs to the guild and sends cloth to be finished by outside labor to have at least half the work done by fellow members of his guild.

The ordinance is a sign that the Merchant Tailors Company had lost control of the cloth-finishing trade. Its members, according to the charter of 1502, were permitted to practice any art connected with the making of men's apparel,[26] but the members of the Clothworkers Guild, incorporated

[23] Year-Book, 40 Edw. III, Pasch., pl. 8 (ff. 17–18). [24] Moore *576 (K.B., 1599).
[25] 1 Clode, Early History of the Guild of Merchant Taylors 393–94 (1888).
[26] Ibid., at 198.

about 1530,[27] had overlapping rights. It therefore became a constant question whether any fuller or shearman was to belong to the Tailors or Clothworkers, whether his work was to be approved, his prices set, and the number of his apprentices regulated by the one company or the other. The dispute led to rioting in the streets of London, litigation, and requests that Parliament settle the matter by legislation.[28] It led also to the Tailors' by-law, the need for which reflected the Tailors' weakness and also the decay of the entire system of guild regulation. The system presumed that each trade would be regulated by its own guild, but with increasing specialization of labor it became more and more difficult to define the limits of a trade or to keep the guilds from splitting into smaller units. There were disputes within disputes: at the very moment that the Tailors were fighting the Clothworkers, the fullers and shearmen who made up the latter company were quarreling over the power to set the price of "rowing"—the fullers insisting that it should be theirs alone, while the shearmen claimed it as theirs by custom.[29] The conflicting interests of related but distinct trades led the Feltmakers to separate from the Haberdashers, and the Glovers from the Leather-sellers. Competition between the guilds led fourteen smaller guilds to petition the government of London to restore the old system whereby "in ancient times the company of artificers or handicraftsmen of the city had reserved the *only* use, trade, or exercise of their several arts and handicrafts,"[30] but the petition was never granted because the Aldermen of London could not restore the economic conditions that had made the guild system possible. One by-product of this general decay was the ordinance in which the Tailors, with a disarming appearance of fairness, decreed that half of the cloth-finishing done for its members must be done by its members.

In *Davenant v. Hurdis*,[31] the by-law was tested. Davenant, a cloth-merchant of the Tailors' company, had sent out twenty cloths to be finished, but refused to give an equal number to members of his guild, and was assessed a fine of 10 shillings per cloth, as the by-law provided. He refused to pay, whereupon the Company instructed Hurdis, its beadle,[32] to take from Davenant goods equal in value to the fine. Davenant brought

[27] Unwin, Industrial Organization in the 16th and 17th Centuries 40–45 (1904).

[28] Clode, op. cit. supra note 25, at 199–203.

[29] Clothworkers' Court Book, April 8, 1567, in Unwin, op. cit. supra note 27, at 231.

[30] Unwin, The Gilds and Companies of London 262 (1938 ed.).

[31] Moore *576 (K.B., 1599).

[32] Clode, op. cit. supra note 25, at 81.

an action of trespass, and Edward Coke, then Attorney General, appeared for him. Coke questioned the authority of the Tailors to make such a by-law or to distrain for the fine, but his principal arguments were that the by-law was unreasonable, and contrary to law. It was unreasonable, he maintained, because it absolutely required the merchants of the company to give their business to the clothworkers, but did not require the latter to provide quick service, good workmanship, or reasonable prices for this business; as a result the merchants might be "utterly impoverished and forced to deceive their customers."[33] It was illegal because it made a monopoly: the same authority that gave the Tailors' Company power to make by-laws keeping half of cloth-dressing to their members would justify them in gradually appropriating the whole of the trade to their own sole use, until finally there would be no cloth-dressing except at their pleasure, and all other clothworkers would be unemployed and live on relief.[34] Coke concluded that a by-law which if extended would give such monopoly powers and bring about such results must be against the public good, and cited precedents to prove that it must therefore be void. But the curious collection of authorities to which he appealed demonstrates how difficult it was to find a traditional basis in common law for the position he was taking. He could merely cite a number of cases in which by-laws or patents were held valid because they were for the public good: a regulation that all ships must harbor in one port and no other, a grant by the King giving a skilled foreigner the sole right to make sailing canvas, and another giving a skilled projector exclusive right to drain lands, a by-law that all cloth sold in London must first be inspected and passed at Blackwell Hall, a by-law of St. Albans requiring each inhabitant to pay a contribution toward cleaning the town, and by-laws for the maintenance of bridges, walls, and similar public works. From these instances, Coke concluded: "but by-laws that establish monopolies are against common law and void." Yet the only direct authority that he offered for this rule was a text from the Civil Law,[35] though he himself maintained that the Civil Law was not authoritative in English courts.[36]

It was against just this weakness in Coke's argument that Francis Moore, attorney for Hurdis, made his principal attack. He conceded that laws ought to be for the public good and that the by-law in question

[33] Moore *576, *580–81 (K.B., 1599).

[34] Ibid., at *579–80.

[35] Ibid., at *580.

[36] 2 Coke, Institutes 98 (1797). See Wagner, The Common Law and Free Enterprise: An Early Case of Monopoly, 7 Econ. Hist. Rev. 217, 218 (1936).

would be void if it created a monopoly. But he denied that it did so, for the by-law did not prohibit any clothworker from using his trade since it regulated the disposal of only a fraction of the business. Moreover, he continued, "if this by-law were really a monopoly, then all the privileges and customs of cities and boroughs, tending to exclude foreigners and to give the sole trading within the city or borough to its own freemen, could be called monopolies and illegal; from which would ensue the decay of all cities and boroughs in the realm . . . which until this day have never been disallowed as monopolies against law and common right."[37] This argument, telling as it was against Coke's assertion that restrictive ordinances were bad at law, failed to convince the judges, who unanimously held that "a rule of such nature as to bring all trade or traffic into the hands of one company, or one person, and to exclude all others, is illegal."[38]

The decision represented an innovation in the law as much as in economic policy. There is no reported common-law case on monopoly prior to *Davenant v. Hurdis;* Coke later mentioned in Parliament some unreported cases,[39] but their precise content is unknown. The willingness of Francis Moore, Hurdis' attorney, to concede that a monopoly would be void at common law does not necessarily indicate that the legal principle was well-established; it may, rather, show the intensity of public opposition to monopolies, in which Moore shared.[40] A number of prior cases are known, but these were heard in the Star Chamber, Privy Council, and other prerogative courts, which generally defended such monopolies as proper exercises of the King's power.[41] The law was still so divided on the validity of monopolies as late as 1624 that Parliament felt it necessary to include in the Statute of Monopolies a provision that "all monopolies . . . and the force and validity of them and of every of them, ought to be and shall be forever hereafter examined, heard, tried and determined by and according to the common laws of this Realm and not otherwise."[42]

[37] Moore *576, *587 (K.B., 1599).

[38] Ibid., at *591: "prescription de tiel nature de inducer sole trade. . . ." "Prescription" does not seem to be used here in its specific technical sense.

[39] 1 House of Commons Journal 555 (March 15, 1621), 606 (May 3, 1621). Fox, op. cit. supra note 22, at 119, cites three cases before Davenant v. Hurdis "in which monopoly grants were considered by the courts of common law." Of these, the case of John the Dyer did not concern a grant of monopoly (see pages 373–74 infra) and Hasting's case, Noy *182, and Humphrey's case, Noy *183, [both mentioned in Darcy v. Allen, Moore *673 (1603)], were tried in the Exchequer.

[40] In 1597 Moore introduced a motion in Parliament against monopolies and was chairman of the committee to which the motion was referred. 2 Cheyney, A History of England from the Defeat of the Armada . . . 296 (1926). He also participated in the debates of 1601. 2 Tudor Economic Documents 274 (1924).

[41] Fox, op. cit. supra note 22, at 119 et seq.

[42] 20 Jac. I, c. 3 (1624).

The next step, and perhaps the greatest single one, in creating the modern common law on monopolies was *Darcy v. Allen*, or *The Case of Monopolies*,[43] decided in 1603. Where *Davenant v. Hurdis* established that a corporate by-law was invalid if it created a monopoly, *Darcy v. Allen* went further, and laid down the principle that even a royal grant by patent would be invalid if it did so. Queen Elizabeth granted Darcy, her groom, a patent for a monopoly of the manufacture and importing of playing cards.[44] In 1601, soon after Elizabeth issued her proclamation on monopolies, Allen, a London haberdasher, made and sold some playing cards, and Darcy brought an action of infringement. The Court of King's Bench unanimously held the patent void.

They held it void as a "dangerous" and "unprecedented" innovation, apparently because no other patent of this sort had previously been issued under the Great Seal. They held it void although it had undoubtedly been granted by the Queen, but in order not to attack royal prerogative directly, they adopted the fiction that "[t]he Queen was deceived in her grant; for the Queen, as by the preamble appears, intended it to be for the weal public, and it will be employed for the private gain of the patentee, and for the prejudice of the weal public." It prejudiced the public good by raising the price and lowering the quality of playing cards, but even more by depriving various workmen of a living. In explaining this main objection, the court said that

> All trades, as well mechanical as others, which prevent idleness (the bane of the commonwealth) and exercise men and youth in labour for the maintenance of themselves and their families, and for the increase of their substance, to serve the Queen when occasion shall require, are profitable for the commonwealth and therefore the grant to the plaintiff to have the sole making of them is against the common law and the benefit and liberty of the subject.[45]

In short, Darcy's patent was held void on the argument that it violated the right of others to carry on their trade.

If the common law recognized each man's right to work at a lawful trade, as the courts of this period became fond of asserting, that right was neither simple nor absolute. Its basis was the feeling that a man should not be denied the means to earn a living: he and his family ought not to starve, his neighbors ought not to be burdened by supporting him, and the

[43] 11 Co. Rep. *84, Moore *671 (K.B., 1599), Noy *173. These reports are collated in Gordon, op. cit. supra note 20, at 193–232.

[44] The patent was originally awarded in 1578, passed through several hands, and was granted to Darcy in 1598. Cheyney, op. cit. supra note 40, at 307–8.

[45] Gordon, op. cit. supra note 20, at 226.

Crown should not be deprived of his contribution to the nation's wealth and power. This right to work was defended by statute and proclamation against foreign competition. A typical statute of this sort, the "Act against Strangers Artificers," passed in 1484,[46] recited the complaint of certain English craftsmen that they were "greatly empoverished" and "likely in short time to be utterly undone for lack of occupation" because of foreign competition, and proceeded to limit importation of certain goods. This sort of protection of domestic workmen was enforced before this time,[47] and after: even toward the end of the seventeenth century, royal proclamations were issued to prevent the importation of rope, hats, knives, gloves, locks, and paper.[48] The common-law right to work was predicated on an economic system that would protect the established trades from competition, whether from foreign workmen, improperly qualified English workmen, overly aggressive guilds, or domestic monopolists. The right to work was protected by giving each guild a monopoly, and Darcy's grant was condemned not because it was a monopoly and therefore necessarily bad, but because it was a bad monopoly.

While the law prior to the eighteenth century supported every man's right to follow his trade, it also strictly limited and regulated this right. The nature of such controls is well illustrated in the third leading case on monopolies decided before the Statute of 1624, the *Ipswich Tailors'* case of 1614.[49] The tailors' guild of Ipswich had a by-law forbidding anyone from practicing his trade in the town unless he had served his apprenticeship under the Company or had been given its approval. They brought suit against one Sheninge for breaking this rule, but the court held that the by-law was invalid, because "at the common law, no man could be prohibited from working in any lawful trade. . . ." In order to reach so broad a conclusion, the court must have closed its eyes to a series of customs and statutes of great age. The right to follow any lawful trade was qualified, for one thing, by the need to have served an apprenticeship—this condition was imposed not only by guild regulations, dating as far back as the thirteenth century in some cases,[50] but also by the Statute of Artificers of 1562.[51] That statute had not lost its force by 1614, and despite the adverse

[46] 1 Ric. III, c. 12 (1483).

[47] E.g., 18 Hen. VI, c. 4 (1439), and Letter Book G, 130 (Feb. 23, 1362).

[48] Charles II, Proclamations (Nov. 20, 1661); ibid. (Feb. 20, 1675); James II, Proclamations (Apr. 29, 1687); ibid. (Aug. 14, 1687); Steele, Tudor and Stuart Proclamations (1910).

[49] 11 Co. Rep. *53, Godbolt *252 (1614).

[50] Cf., e.g., early charters printed in Consitt, 1 The London Weavers' Company (1933).

[51] 5 Eliz., c. 4 (1562).

decision in the *Ipswich Tailors'* and subsequent cases, it was still followed at the end of the seventeenth century.[52]

The right to follow a trade was limited also by the rule that no man might work at several trades simultaneously. The validity of this rule was argued in the *Ipswich Tailors'* case, for the guild maintained that it rightly refused to approve Sheninge since he was already doing another kind of work. The court apparently decided that the common law did not prohibit this; it seems to have accepted Coke's statement that the prohibition was first introduced by a statute of 1363, and that it was found so harmful that it was repealed in the following year.[53] But Coke was mistaken. The law of 1363 to which he referred[54] ordained, among other things, that each merchant should deal in only one sort of merchandise[55] and "that Artificers, Handicraft People, hold them every one to one Mystery."[56] It may be true, as Coke says, that before this time the common law did not require each artisan to keep to his own trade. It is quite certain, however, that the prohibition was not repealed in the following year. Only the section which directed merchants to restrict their trade to one commodity was repealed;[57] it would have been difficult to restrain a merchant who carried wool abroad from returning with wine, iron, or wax. The section which confined each workman to a single trade stayed in the books two hundred years more, and was indeed reinforced from time to time by specific acts such as that which forbade tanners to be shoemakers or shoemakers to be tanners.[58] The legal principle on which the Ipswich Tailors depended and which Coke denied was not just a momentary aberration from a long-standing common-law tradition. The fact is that the monopolistic powers of guilds, which Coke insisted repeatedly were always void at common law, had really been supported by law. That support first

[52] Francis Kiderby was indicted under the statute in 1669, for setting up as a draper without having been apprenticed. He petitioned the Privy Council that the Crown might drop the prosecution, for, he said, "the Statute though not repealed yet, has been by most of the judges looked upon as inconvenient to Trade and to Encrease of Inventions." Nevertheless he felt sure that a common law court would find him guilty. His petition was granted. Privy Council Register (Oct. 29 and Dec. 17, 1669), quoted by Unwin, op. cit. supra note 27, at 252. Cf. Wade v. Ripton, 2 Keble *125, Siderf. *303 (1666).

[53] 11 Co. Rep. *53, *54 (1614).

[54] 37 Edw. III (1363).

[55] Ibid., at c. 5. [56] Ibid., at c. 6.

[57] 38 Edw. III, c. 2 (1364), repealed 37 Edw. III, c. 5 (1363).

[58] 37 Edw. III, c. 6 was repealed by 5 Eliz., c. 4 (1564). The Act on tanners and shoemakers was 13 Ric. II, s. 1. c. 12 (1389).

began to be withdrawn in the beginning of the seventeenth century,[59] under the pressure of, among other things, Coke's powerful but inaccurate polemics.

There is no doubt that the series of cases at the turn of the seventeenth century radically changed the attitude of the common law toward monopolies But it must be borne in mind that this change was also a consequence of the decay of the monopolistic system from within. Pecche was not attacked by irate consumers for raising his prices, but by irate subjects objecting to an unpopular minister. Similarly, *Davenant v. Hurdis* was not a dispute between a freedom-loving tradesman and a tyrannical guild as much as a conflict between two guilds for control of an industry. And *Darcy v. Allen* was not the action of a solitary champion bravely contesting the monopoly of a powerful courtier; it has been shown instead that Allen was supported in the case by the Mayor and Aldermen of London, who, regarding Darcy's patent as an attack on all the trades and privileges of the City, "comforted and animated [Allen] to continue his selling of cards" and promised to pay the costs of any legal action that might follow. When Allen submitted a bill for his costs in defending himself against Darcy, the Mayor refused to pay, but Allen sued him and recovered.[60]

Moreover, the mercantilist system of private and corporate monopolies, though very much weakened by 1600, was still too widespread to be destroyed by the application of common-law remedies in specific cases. It was seriously limited, and in the end destroyed, by legislation. The first important law contributing to that result was the Statute of Monopolies of 1624, which, however, has a deceptive ring. For though it was certainly directed against monopolies, it was based not on a preference for competition, but on constitutional objections to the power which the Crown presumed in granting monopolies and to the arbitrary reasons for which it had granted them. Parliament did not at this period oppose monopolies in themselves. As Bacon told the House of Commons in 1601, its attitude was inconsistent and suspect:

> If her Majesty make a patent or a monopoly unto any of her servants, that we must go and cry out against: but if she grant it to a number of burgesses or a corporation, that must stand, and that forsooth is no monopoly.[61]

[59] They were still supported, for instance, in The Warden and Corporation of Weavers in London v. Brown, Cro. Eliz. *803 (1600), where the court held that Brown did not come under the weavers' control, for though he sold his goods in London, he wove them elsewhere; but the court added that this judgment did not question the guild's right to control weaving in London: "It were a good custom . . . being used time [immemorial]." Ibid., at 803.

[60] Davies, Further Light on the Case of Monopolies, 48 L.Q. Rev. 394 (1932).

[61] Prothero, Statutes and Constitutional Documents 112 (1906).

This inconsistency the House of Commons carried over into the Statute of Monopolies, the first section of which declared void "all monopolies and all commissions, grants, licenses, charters, and letter patents heretofore made or granted, or hereafter to be made or granted to any person or persons, bodies politic or corporate whatsoever, of or for the sole buying, selling, making, working, or using of anything, or of any other monopolies"; the ninth section nevertheless provides that the Act shall not apply to any cities or towns, or any of their privileges, "or unto any corporations, companies, or fellowships of any trade, occupation, or mystery, or to any companies or societies of merchants within this Realm, erected for the maintenance, enlargement, or ordering of any trade of merchandise. . . ." And this inconsistency, which symbolized Parliament's willingness to have monopolies, provided Parliament alone granted them, was not merely a matter of words in a statute. It justified the final irony in the case of *Darcy v. Allen:* only a few years after Darcy's monopoly of playing cards was judged void at common law, the same monopoly was given, under authority of the Statute of Monopolies, to the Company of Card Makers.[62]

The Statute of Monopolies soon put an end to the arbitrary granting of private monopolies. But it was not intended to abolish customary monopoly privileges of corporations. Cities and boroughs, guilds, and chartered trading companies continued to exercise their monopoly powers to exclude strangers from various trades.[63] The common law continued to protect them, though with lessening fervor as the influence of economic liberalism grew, and some of these monopolistic controls were finally abolished only by legislation in the nineteenth century.[64]

II

Throughout these early monopoly cases the complaint is made that practices are objectionable because they tend to raise prices. But even this complaint did not arise from opposition to monopolies. It did not mean that the common law early in the seventeenth century favored competition or endorsed the determination of prices by the free play of the market. The common law favored "low" prices rather than free prices, and accepted as a matter of course that all important prices would be set by political or corporate authorities. The complaint meant only that Englishmen objected to private efforts to raise prices, and that they readily attributed a rise in prices to the evil machinations of profiteers. This super-

[62] Fox, op. cit. supra note 22, at 128 n. 21.

[63] Cases are voluminously noted in 32 Halsbury's Laws of England 345 n. 'o' (Hailsham's 2d ed., 1939).

[64] Ibid., and see page 375 infra.

stition was written into the early common law in the form of provisions against forestalling, regrating, and engrossing.

The body of law concerning these crimes has been thought to be an integral part of the law on monopolies because forestalling and the associated offenses seem at first sight to be older names for the modern monopolistic tactic known as "cornering the market"; and because since the seventeenth century, "engrossing" has become almost synonymous with "monopolizing." Jeffreys coupled the terms in this way when he gave his opinion in *The East India Co. v. Sandys*,[65] and so did the authors of the Sherman Act when they explained the meaning of the word "monopolize" in the second section of that Act.[66] But in fact the two bodies of law are quite distinct: they evolved from separate statutes, one did and the other did not raise questions of royal prerogative, and whereas the modern law on monopoly by patent was laid down early in the seventeenth century, that on forestalling did not take its present shape until almost two hundred years later. The basic legal difference is that the monopolist had a legal warrant for his activity, whereas the forestaller was justified by no custom, grant, or statute whatsoever.

Contrary to monopolies by patent, which always were and still are legal —the principal changes being in who gives and who may receive them— forestalling was always illegal, and ceased to be so only when the crime was altogether abolished. "Forestalling" in the common law before the thirteenth century is said to have been an inclusive term for all unlawful attempts to raise prices.[67] It came to be a more particular term: in the year 1266, the first statute prohibiting it defined forestallers as those "that buy anything before the due hour, or that pass out of the town to meet such things as come to the market."[68] "Regrating" meant simply retailing, buying in bulk and selling in small lots, and "engrossing," in its original narrow meaning, was to buy crops in the field before they were harvested or at least before they were ready to come to market. These offenses were indictable at common law, and various statutes assigned punishments ranging, according to the temper of the time, from fines and forfeiture to banishment and even death.[69] Such statutes were passed periodically from

[65] 10 Howell's State Trials 372, 538 (1685): "though the word Monopoly, or Engrossing. . . ."

[66] 21 Cong. Rec. 3152 (1890): "monopoly . . . is the sole engrossing to a man's self. . . ."

[67] Illingworth, An Inquiry into the Laws . . . Respecting Forestalling . . . 14 (1800).

[68] 51 Hen. III, § 6 (1266).

[69] Banishment from the town where forestalling was committed was imposed by a statute of uncertain date, probably prior to 1327; 1 Statutes of the Realm 197, 202 et seq. [But cf. Winfield, Chief Sources of English Legal History 93 (1925).] Forfeiture was imposed by 25

the thirteenth until the late seventeenth centuries, in some cases against the forestalling of certain specified commodities, in others against forestalling generally. But the application of these general statutes, as of the common law itself, was relatively narrow, and usually only the forestalling of food—or more precisely, of "dead victuals"—constituted a crime. One statute forbade engrossing of hides and oak-bark,[70] and another, of cloth;[71] but the fact that Parliament legislated for these commodities in particular argues that it did not consider them to be included under the general laws against forestalling. In a few scattered cases, also, courts found defendants guilty of forestalling land or houses,[72] but here "forestalling" could only be brought in by stretching an analogy, for the offense was generally understood quite literally as buying commodities before they had been carried into the actual market place or before the market had officially opened. Perhaps the strongest evidence on the point occurs in *Rex v. Waddington*,[73] one of the last important English cases of this type,[74] in which the defendant, having been charged with forestalling and engrossing hops, argued that his was no offense since hops were no victual, and the Court appeared to agree that if it had not been, there would have been no offense.

The major objective of laws against forestalling was to keep food prices low. Such laws fit very neatly into the more general price-fixing program administered by medieval and, later, mercantilist governments. Local authorities of manors, cities, and guilds had customary rights to control food prices; kings issued proclamations and parliaments passed statutes for the same end; all these are implicitly confirmed in a statute of 1533 which gave certain members of the Privy Council as well the right to set "reasonable prices" of "cheese, butter, capons, hens, chickens, and other victuals necessary for man's sustenance."[75] The work of surveillance would be much easier if all sales were made publicly in the market, and so forestalling and engrossing, means of evading the market, were seen as attempts to evade price-controls.

Edw. III, c. 3, § 4 (1350); death by 27 Edw. III, c. 11, § 2 (1353), repealed 38 Edw. III, c. 6, § 1 (1363).

[70] 1 Jac. I, c. 22, §§ 7, 19 (1604).

[71] 1 Phil. & M., c. 7 (1554). Liber Albus 172–73 (comp. 1419, translated by Riley, 1861), is the case of a merchant fined by a London court for forestalling cloth; uncertain date prior to 1419.

[72] Fox, Monopolies and Patents 21 n. (1947). [73] 1 East *143 (1800).

[74] Sanderson, Restraint of Trade in English Law 97 (1926).

[75] 25 Hen. VIII, c. 2, § 1 (1533).

But to maintain low food prices was not the sole objective of the laws against forestalling. Just as monopolies by patent were attacked by those who feared to lose their own monopoly powers, so forestalling was abhorred not only by a public which hated high prices but also by those who saw in it an infringement of their privileges as owners of markets. Rights to hold markets were granted or confirmed by the Crown, and established local but powerful monopolies. What was given was not the mere right to hold a market, but an exclusive right. The extent of such privileges is illustrated in the case of the *Abbot of Westminster*.[76] The Abbot brought an action against one who sold cloth in London; William the Conqueror had given him a patent to hold a fair for thirty days, during which nobody should buy or sell merchandise at any other place within a radius of seven miles. The court presumably upheld the grant, and similar grants were upheld regularly. The owners of markets often had an intense interest in protecting their exclusive rights, for some of them had the right to charge a toll on certain goods sold in the market, and all of them were entitled to charge fees for market stalls put up on their land. To hinder sellers from coming to a market was therefore to deprive the owner of the market of stallage fees; thus the Prior of Coventry, in a suit against several who sold merchandise outside his market, declared that he thereby "lost stallage, terrage and cottage, etc., wrongfully and to his damage."[77]

The objectives of statesmen and the interest of owners of markets coincided with the prejudices of the public. They considered forestalling, engrossing, and regrating the typical tricks of middlemen and speculators, and were convinced that merchants who used such tactics were parasites profiting by the distress of others. They could see nothing but evil and selfishness in such practices. Thus a commission of inquiry into forestalling in Suffolk in 1411 reported: Geoffrey Russell bought sixty quarters of barley at forty pence a quarter and sold them for twice as much; "John Cok and John Joye . . . secretly bought, in private and secret places, sixty quarters of wheat . . . a quarter at eight shillings; whereas in open market, the same was sold for six shillings per quarter, &c, and so the aforesaid John and John are common forestallers of corn"; and "Simon Basket . . . bought at Beccles, Owtehole, and Brompton, and in divers other places, forty quarters of wheat, of the price of six shillings per quarter, and conveyed the same coastwise into divers other parts, whereby the price of a quarter of wheat was raised to ten shillings . . . and so the afore-

[76] Registrum Brevium, f. 107. The case is cited in Darcy v. Allen, 11 Co. Rep. *84, Moore *671 (K.B., 1599), and East India Co. v. Sandys, 10 Howell's State Trials 372, 538 (1685).

[77] Year-Book 2 Edw. II, pl. 141 (1308).

said Simon is a common forestaller."[78] The practices described show no sign of being, properly speaking, monopolistic; they appear on the contrary to have been acts of speculation, arbitrage, or wholesaling; but most men continued to identify the two phenomena until the new economic theory in the eighteenth century taught a few of them at least that speculation was no more profitable to the merchant than to the community at large, and that the community had as much to gain as the merchant from free trade.

The development of laissez-faire economic theory accounted for the abolition of the laws against forestalling. After 1552, when the great declaratory statute against forestalling was passed,[79] the general prohibitions were reasserted periodically when food prices became unusually high. A Commonwealth Parliament passed a law in 1650 denying habeas corpus to defendants in any action concerning the buying or selling of foods,[80] and William III issued a proclamation in 1698 insisting that the laws on forestalling be administered with full force.[81] Another such occasion arose in 1766 when corn prices were particularly high. Many complaints were voiced, tumults and riots took place, "in which, as usual in popular commotions, great irregularities took place," many lives were lost, order was restored only after the militia was called out and a number of rioters sentenced to death and hanged. George III tried to improve things by issuing a proclamation to put in force the statutes against forestallers. But, *The Annual Register* noted, many doubted whether such action could be of any use: "It was apprehended that this measure would have an effect contrary to the intentions of the council, and by frightening dealers from the markets, would increase that scarcity it was designed to remedy."[82]

The doctrine hinted at in this comment, that public regulation of the market would produce worse results than the free action of merchants, was still novel in 1766 but beginning to gain force. It dominated the committee of the House of Commons which reported in the following year "that the several laws relating to Badgers, Engrossers, Forestallers, and Regrators, by preventing the circulation of, and free trade in, corn, and

[78] Plea Roll 12 Hen. IV, 6 (1411), quoted in Illingworth, op. cit. supra note 67, at 240–42. Why Cok and Joye should have committed a crime at high prices rather than trade legally at low must remain a mystery.

[79] 5 & 6 Edw. VI, c. 14 (1552).

[80] Statute of Oct. 23, 1650, Firth and Rait, 2 Acts and Ordinances of the Interregnum 442 et seq. (1911).

[81] William III, Proclamation Oct. 13, 1698.

[82] Annual Register 39–40 (1767).

other provisions, have been the means of raising the price thereof in many parts of this Kingdom."[83] The report suggested that those laws should be abolished, and action of this sort was finally taken in 1772, when a bill was prepared, reported by Edmund Burke—who was the main exponent in Parliament of this measure—and quickly passed by Commons.[84] The Act repealed the various statutes against forestalling because, as its preamble said, "it hath been found by experience that the restraints laid by several statutes upon the dealing in corn, meal, flour, cattle, and sundry other sorts of victuals, by preventing a free trade in the said commodities, have a tendency to discourage the growth, and to inhance the price of the same."[85] With the passage of this Act the cause of free trade seemed to be triumphant; the crime of forestalling had been abolished—so Blackstone, among others, thought[86]—and Adam Smith's remark, published three years later, that to fear forestalling was like fearing witchcraft seemed to be more useful as a contribution to public education than to practical policy.[87]

For a short while, however, the law against forestalling was revitalized by Lord Kenyon's decision in *Rex v. Rusby*.[88] Rusby was indicted in 1799 for regrating thirty quarters of oats, and presumably rested his defense on the Act of 1772. Kenyon held, however, that "though in an evil hour all the statutes which had been existing above a century were at one blow repealed, yet, thank God, the provisions of the common law were not destroyed"[89] and found Rusby guilty at common law. The vigor of Kenyon's address against Rusby so inflamed the public that a mob of Londoners rioted, tried to lynch Rusby, and ended by pulling down his house;[90] the public was apparently not so convinced as Burke, Smith, and Parliament that forestalling was economically beneficial, or, at least, that laws prohibiting it were more harmful than the thing itself. After Kenyon's time, however, there were no further common-law prosecutions against forestalling,[91] and to make quite sure, Parliament in 1844 passed a law repeal-

[83] House of Commons Journal (Apr. 8, 1767).

[84] Ibid. (drafting committee appointed March 13, 1772, bill read May 6, passed May 20). For Burke's views on the subject, see his Thoughts and Details on Scarcity, 5 Works 133, 150 et seq. (Nimmo's ed., 1899).

[85] 12 Geo. III, c. 71 (1772).

[86] 4 Blackstone, Commentaries 159 (5th ed., 1773).

[87] 2 Smith, Wealth of Nations 34–35 (Canaan ed., 1922).

[88] Peake Add. Cas. *189 (1800). [89] Ibid., at *192.

[90] Barnes, History of the English Corn Laws 81–82 (1930).

[91] Sanderson, op. cit. supra note 74, at 98.

ing all the remaining statutes against it, and utterly abolishing the common-law crimes of forestalling, engrossing, and regrating.[92]

Clearly, then, the laws against forestalling and engrossing, which some have tried to identify as a fount of modern antitrust law, did not have the required character. They were of narrow scope, applying almost exclusively to trade in foodstuffs; they were part of a program to regulate all economic activities; like the common law against monopolies by patent, they were supported by monopolists—in this case, the owners of markets—who found them useful protection; and they were finally repealed by the supporters of free trade and in the name of free trade.

III

Because the Statute of Monopolies settled that branch of law into its present narrow concern with patents, and the Acts of 1772 and 1844 altogether did away with the law against forestalling, the only English tradition from which modern antitrust law could grow were the bodies of law against contracts in restraint of trade and combinations in restraint of trade. The manner in which those laws were interpreted during the nineteenth century, however, very much weakened their capacity for controlling modern monopolies.

The common law relating to contracts in restraint of trade stems from the *Case of John Dyer*,[93] decided in 1414. The report of the case is meager. John was sued for breaking his bond not to practice the trade of dyeing in his home town for a half year; he apparently maintained that he had not broken his bond, and seems to have won the case. One of the judges suggested that John could have used the stronger defense of demurring at law, as the condition of the bond was illegal, and he continued in words which have become all too famous in the literature on monopolies: "By God, if the plaintiff were here he should go to prison until he paid a fine to the King." The case was an extremely powerful precedent until the beginning of the eighteenth century.

But for all its legal power, the *Case of John Dyer* does not demonstrate—as so many have believed—that the common law always condemned restraints on trade. If one seventeenth century lawyer could cite the case as authority for that view, his opponent could counter with a much older

[92] 7 & 8 Vic., c. 24 (1884). 4 Holdsworth, History 379 (1924), says that the forestalling laws were repealed by 6 Geo. IV, c. 129 (1825), under the influence of "the economists of the school of Ricardo." There is little evidence that Ricardo or his school particularly affected the passing of the act of 1844, and the act of 1825 which Holdsworth cites has no bearing on forestalling, being instead a combination act. See page 381 infra.

[93] Year-Book 2 Hen. V, 5B (1414).

precedent, the *Case of the Archbishop of York*,[94] in which a court upheld the custom of the Archbishop's manor at Ripon that no one should operate a dyeing-house there without the Archbishop's license. If the common law is supposed to have been such an ardent protector of free trade, why was it prepared to uphold an Archbishop's power to keep a dyer from following his trade in Ripon forever, but not an ordinary man's power under a voluntary agreement to keep a dyer from following his trade in Dale for half a year?

The answer, of course, is that the common law prior to the fifteenth century did not favor free trade but reached its decision in *John Dyer*'s case on quite different grounds. The restraining agreement in the *Dyer*'s case was embodied in a bond, and it has been suggested that judges distrusted bonds because they were so often oppressive, and lost few opportunities to hold them void.[95] Lord Macclesfield, in his famous opinion in *Mitchel v. Reynolds*,[96] held that the issue was not so much whether the restraining agreement was by bond or by contract but whether it was based on a good and adequate consideration. Yet, as he understood *John Dyer*'s case, the fact that it rested on a bond seemed material; "for suppose," he wrote,

> (As that case seems to be) a poor weaver [sic], having just met with a great loss, should, in a fit of passion and concern, be exclaiming against his trade, and declare, that he would not follow it any more, etc., at which instant, some designing fellow should work him up to such a pitch, as, for a trifling matter, to give a bond not to work at it again, and afterwards, when the necessities of his family, and the cries of his children, send him to the loom, should take advantage of the forfeiture, and put the bond in suit; I must own, I think this such a piece of villainy, as is hard to find a name for. . . .[97]

Lord Macclesfield evidently thought that no court should uphold an agreement made in such circumstances, and perhaps earlier common-law judges did too.

But the more important basis for deciding against the restraint in the *Dyer*'s case was the principle, so important in the cases on monopolies by patent, of the individual's right to work. It may appear that to prevent a man from following one trade in one particular town for six months did not very seriously limit his right to work: he might take up another trade, or

[94] Reg. Brev., f. 105. The two cases were confronted in Darcy v. Allen, op. cit. supra note 43.

[95] Sanderson, op. cit. supra note 74, at 14. Cf. Clerk v. Taylors of Exeter, 3 Lev. 241 (1685), in which the Exchequer held that in all the previous cases on restraint of trade the agreement had been disapproved if by bond, approved if by contract (assumpsit).

[96] 1 Peere Wms. *181 (1711).

[97] Ibid., at 193.

move to another town. But in the fifteenth century, those alternatives were not in fact open to him. In order to take up another trade, he would have had to pass through an apprenticeship of seven years, or with great difficulty and expense satisfy a guild that he was a master of its craft, and this remedy was for all practical purposes ruled out. Nor could he more readily practice his own trade in a new town, for the guilds and municipal corporations of each place had by-laws to prevent strangers—that is, anyone not free of the town or its guilds—from entering into competition with citizens.[98] The whole guild system, therefore, made it nearly impossible for a tradesman to earn his living if he did not practice his own trade in his own town, and this was the main reason why, as long as the guilds maintained their power, contracts in restraint of trade were held void. To have done otherwise would have been to concur in arrangements by which men deprived themselves of their means of support. The power of the guilds to regulate entrance into trade had begun to weaken by the sixteenth century, although cities and towns retained fragments of such powers until 1835, when the Municipal Corporations Act finally gave anyone the right to keep any shop or follow any trade in any borough.[99] It was with the decay of this power in the guilds that the law on contracts in restraint of trade came to change, and the time when the courts refused to uphold the restrictive powers of the Ipswich Tailor's guild was also the time when they first held valid a contract in restraint of trade.[100]

From then on, the law on the subject became more and more complex, since each case involved two contrary principles. On the one hand, the common law was inclined to uphold contracts in restraint of trade for the same reasons which moved it to sustain any good contract. To own property implied the right to dispose of property by contract, and if a reasonable man disposed of his property in a way which he considered good, it was not for the court to tell him he was mistaken.[101] On the other hand, the

[98] Cf. Sanderson, op. cit. supra note 74, at 15.

[99] 5 & 6 Wm. IV, c. 76 (1835). London was not considered a borough under the Act, and cases to exclude "foreigners" from certain employments in London continued until later in the nineteenth century. 32 Halsbury's Laws 345 (1939).

[100] Ipswich Tailors' case, 11 Co. Rep. *53 (1613); Rogers v. Parrey, 2 Bulst. *136 (1614). In the earlier cases, the restraints were held void, John Dyer's case being cited as authority for each: Anon., Moore *115 (K.B., 1578)—a mercer's apprentice bound himself not to exercise his craft for four years in Nottingham; Anon., Moore *242 (K.B., 1587), 2 Leonard *210—a blacksmith bound himself not to practice the trade in South-mims; Colgate v. Bacheler, Cro. Eliz. *872, Owen *143 (1601)—a haberdasher gave bond not to trade in Canterbury or Rochester for four years, of which Anderson, J. said that "he might as well bind himself, that he would not go to Church."

[101] Jollife v. Broad, 2 Roll. Rep. 201, Cro. Jac. *596, 1 Wm. Jones 13 (1620): Restraint ancillary to sale of a mercer's business in Newport held void in Common Pleas, reversed in

common law was inclined to invalidate contracts in restraint of trade because they deprived a man of the means to earn a livelihood,[102] or because they deprived the public of the advantages of competition.[103] The first of these reasons prevailed until the eighteenth century, the second slowly replaced it. The conflict of these principles, and their application to the particular circumstances of each case, have resulted in the general rule, still true today, that some contracts in restraint of trade are good and others are bad. The basis on which the distinction should be made was first formally stated in *Mitchel v. Reynolds*, decided in 1711; it was stated in broader terms—or as some think, changed—by the decision in *Nordenfelt v. Nordenfelt* in 1894.

In *Mitchel v. Reynolds*[104] the defendant assigned to the plaintiff the lease of a bakery in a certain parish of London for five years, and undertook to pay the plaintiff £50 damages if he should work as a baker within that parish during those five years. The plaintiff brought suit for the damages, and though the defendant pleaded that since he had served his apprenticeship as a baker and had been admitted to the guild no private person could lawfully prevent him from working at that trade, Chief Justice Parker (later Lord Macclesfield) found for the plaintiff. The fame and great interest of this case is not in the decision, but in the opinion, for there Lord Macclesfield very systematically classified all restraints of trade and arrived at his long-lasting rule for distinguishing good restraints from bad. He first divided all restraints of trade into involuntary or voluntary; the contract in issue was clearly a voluntary restraint. Among voluntary restraints he distinguished between those "where the restraint is general not to exercise a trade throughout the kingdom, and where it is limited to a particular place." General restraints, he held, had always been held void, and should be, "being of no benefit to either party, and only oppressive."[105] But particular restraints are of two sorts: without consideration, "all of which are void by what sort of contract soever created," and Macclesfield here cited the *Dyer*'s case as evidence for this point; and with consideration. Macclesfield's rule, therefore, was that a contract in re-

---

King's Bench, affirmed in Exchequer. Dodderidge, J. in K.B.: "It is the usual course of men in their old age to turn over their trade to another. . . ."; 2 Roll. Rep. 201, 203. Similarly, Bragge v. Stanner, Palmer *172 (1621); Pragnell v. Goff, Style *111, Aleyn *67 (1648); Hunlocke v. Blacklowe, 2 Wms. Saunders *156, 2 Keble *674 (1670).

[102] See the cases listed at note 100 supra. Also Ferby v. Arrosmyth, 2 Keble *377 (1668).

[103] See the cases listed at note 108 infra.

[104] 1 Peere Wms. *181 (1711).

[105] Ibid., at *182.

straint of trade can be good only if the restraint is particular and the contract "appears to be made upon a good and adequate consideration, so as to make it a proper and useful contract."[106] Lord Macclesfield had discovered this much from precedent, but being a lawyer of his time, he was interested to show not only what the law was, but also the reasonable explanation of why it should be so, and therefore he added:

> [T]he true reasons of the distinction upon which the judgments in these cases of voluntary restraints are founded are, 1st, the mischief which may arise from them, 1st, to the party, by the loss of his livelihood, and the subsistence of his family; 2dly, to the public, by depriving it of an useful member.[107]

These mischiefs might arise from general restraints but could not from particular restraints; the latter were good because the parties might gain and the public lose nothing from them. Once a man could reasonably be expected to enter a new trade if he sold his last one, or to move to a new place if he bound himself not to trade in his former one, the law became quite willing to uphold fair contracts in which he bound himself to desist from competition within a limited area. And so the test became whether the contract was fair and the area to which it applied limited.

For at least a century, Macclesfield's rule was followed quite religiously. But then new problems in the interpretation of the rule began to enter: How important a consideration was public policy; if a contract did not injure either of the parties, might it still be bad because it interfered with competition? How large an area was "particular" rather than "general"; if the business of one party extended throughout the kingdom and real protection from competition could only be had by a "general" restraint, was a contract of this sort nevertheless to be held bad automatically?[108] How was the word "reasonable," recurring so often in Macclesfield's opinion, to be interpreted; was the real test of a restraint whether it was reasonable, and was Macclesfield only enunciating a special case of the rule, valid for his time but not for always, when he held that particularity and consideration made a restraint reasonable and therefore good? All these questions were decided one way or another by judges during the nineteenth century and finally settled by the decision of the House of Lords in *Nordenfelt v. Maxim Nordenfelt Guns and Ammunition Co.*,[109] and particularly by the rule laid down by Lord Macnaghten.

[106] Ibid., at *185–86.

[107] Ibid., at *190.

[108] See, e.g., Wickens v. Evans, 3 Y. & J. *318 (1829); Wallis v. Day, 2 M. & W. *273 (1837); Mallan v. May, 11 M. & W. *652 (1843); Shrewsbury and Birmingham Railway Co. v. London and North-Western Railway Co., 21 L.J.Q.B. 89 (1851); and Tallis v. Tallis, 1 El. & Bl. *391 (1853).

[109] [1894] App. Cas. 535.

The case arose on a contract whereby Thorsten Nordenfelt agreed not to engage in the ammunition and armaments business, except in behalf of the Company; the restraint to apply for twenty-five years and in all countries. The Company brought action to enforce the covenant by injunction, and Nordenfelt successfully defended himself in the lower courts by arguing that the restraint was general and therefore void. The Court of Appeals reversed the lower court[110] and Nordenfelt then appealed to the House of Lords. That body unanimously held that the restraint, though general, was valid because it was reasonable. Lord Macnaghten, in his concurring opinion, set down what has since been the governing rule in such cases; it very neatly coordinates the considerations of public policy, generality of restraint, and reasonableness:

> The public have an interest in every person's carrying on his trade freely: so has the individual. All interference with individual liberty of action in trading, and all restraints of trade of themselves, if there is nothing more, are contrary to public policy, and therefore void. That is the general rule. But there are exceptions: restraints of trade and interference with individual liberty of action may be justified by the special circumstances of a particular case. It is a sufficient justification, and indeed it is the only justification, if the restriction is reasonable—reasonable, that is, in reference to the interests of the parties concerned and reasonable in reference to the interests of the public, so framed and so guarded as to afford adequate protection to the party in whose favour it is imposed, while at the same time it is in no way injurious to the public.[111]

Only the greatest optimism could have made it appear to Lord Macnaghten that contracts in restraint of trade are at all likely "to afford adequate protection" to a party and at the same time to be "in no way injurious to the public." What makes this entire area of law so difficult is precisely that the interests of the restraining party and of the public are so often opposite. This problem indeed seems to be at the core of *Nordenfelt v. Nordenfelt*, although the court very tactfully covered it over. Vague references were made to the public interest, which was said not to favor or require Thorsten Nordenfelt's competition with the Company. On the other hand, a great deal was said about the Company's need for adequate protection. The Lord Chancellor took notice of the improved means of communication that had become available since *Mitchel v. Reynolds*, and argued that in these new conditions reasonable protection may mean worldwide protection; he decided that because the Company sold its armaments mainly to governments, considering "the nature of the business and the limited number of customers,"[112] it needed protection of such width.

[110] [1893] 1 Ch. 630.

[111] [1894] App. Cas. 535, 565.

[112] Ibid., at 548–50.

And Lord Watson, in his concurring opinion, made it quite clear why it was possible so readily to identify the public interest and the private interest of the covenantee:

> [But] it must not be forgotten that the community has a material interest in maintaining the rules of fair dealing between man and man. It suffers far greater injury from the infraction of these rules than from contracts in restraint of trade.[113]

This remark is the clearest possible indication that by 1894 English law on contracts in restraint of trade was not in any important respect an instrument for the maintenance of a competitive economic order. If ever, then only for a very short period after *Mitchel v. Reynolds* did the courts give the public policy of promoting competition an important part in deciding cases on contracts in restraint of trade. The decision itself in *Nordenfelt v. Nordenfelt*, the words in which Lord Macnaghten expressed his rule of reasonableness, and the dictum of Lord Watson, all declared that competition was no longer public policy, or at least that freedom of contract had become a more important end than freedom of trade.[114]

IV

The law on combinations in restraint of trade was by the end of the nineteenth century narrow and ineffective. Developments both in statute and common law joined to produce this result. The statute law governing combinations became increasingly lenient during the nineteenth century, in response to greater sympathy, abstract as well as sentimental, for the labor unions. The common law, influenced by a feeling that employers should not be denied rights granted to workers, matched the new legal power of the latter with a solicitous concern for employers' combinations; in the end it came to put a higher value on the freedom of entrepreneurs to use any means short of violence to outstrip competitors than on the right of the public to enjoy the advantages of competition.

Legislation governing wages and conditions of labor began with the Ordinance of Labourers passed in 1349,[115] which was confirmed and extended by numerous later statutes. In the sixteenth century, these occasional laws were consolidated in the great Elizabethan Statute of Artificers,[116] which governed apprenticeship and wages, and in the Act of 1548,[117] which provided criminal penalties against any workmen who con-

[113] Ibid., at 552.

[114] Cf. Morris v. Saxelby, [1916] 1 App. Cas. 688, 699.

[115] 23 Edw. III (1349).

[116] 5 Eliz., c. 4 (1562).

[117] 2 & 3 Edw. VI, c. 15 (1548), confirmed by 22 & 23 Car. II, c. 19 (1670).

spired or agreed to raise wages or reduce hours of labor. But there was no thought in these statutes of making it possible for workmen to compete; on the contrary, the sixteenth century legislators who passed these laws to fix the terms and wages of labor hoped to recapture the economic stability that had been shaken by the Black Death, the movement of men from manors to towns, and the early industrial revolution.[118] They wanted competition no more than they understood it. Although these laws, like others of the time, appear, in the light of later developments and interpretation, to express antagonism to monopolistic arrangements and approval of competition, they were really intended to reinforce the system of direct economic control.

The earliest instances of general labor legislation influenced by free-trade theories were the Combination Acts of 1799,[119] which prohibited combinations of workmen only, and the Combination Act of 1800,[120] which superseded it and prohibited masters as well as workmen from combining. It has been said that these Acts were prosecuted more severely against combinations of laborers than against those of masters; it appears that they were not very effective at all; it is certain that they did not settle the problem of changing from a legally regulated labor market to a free one.[121] Riots and violence continued as frequent incidents of labor relations after 1800, and Place, Joseph Hume, McCulloch, and other radical followers of Bentham argued that this disorder would not cease until workers were given a legal right to combine.[122] The Benthamites, besides, advocated a statute permitting workers to combine because this law was implied by two of their basic political principles. These, as Dicey has expressed them, were "the belief that trade in labour ought to be as free as any other kind of trade," and "the well-grounded conviction that there ought to be one and the same law for men as for masters; Adam Smith had, about fifty years earlier, pointed out that trade combinations on the part of workmen were blamed and punished, whilst trade combinations on the part of masters were neither punished nor indeed noticed."[123] In the early decades of the nineteenth century, those who favored freedom of trade tried to achieve their other aim of equality before law by giving workers as well as masters freedom to associate. By rejecting the alternative that would have

[118] Nef, 1 Rise of the British Coal Industry 165–89 (1932).

[119] 38 Geo. III, c. 81 (1799).

[120] Ibid., at c. 106 (1800).

[121] George, The Combination Laws, 6 Econ. Hist. Rev. 172 (1935).

[122] Robbins, The Theory of Economic Policy 106 et seq. (1952).

[123] Dicey, Law and Opinion in England 196 (1905 ed.).

achieved both their objectives, by not denying the right of combination to workers and masters alike, they wasted an opportunity to secure economic freedom against attack by monopolies of either.

But in any case, so strong was the Benthamites' belief in freedom of association, so complete their inability to foresee the monopolistic position which labor unions and industrial combinations would in time achieve, so great their faith that if workers were allowed to combine in friendly associations they would desist from all violence and monopolistic activities, that in 1824 they successfully urged Parliament to pass the Act that gave combinations of workmen and masters alike immunity from all statutory and common-law prohibitions.[124] Workmen proceeded to make the most uninhibited use of that immunity. In the next session of Parliament, Huskisson brought in a bill that reversed the terms of the Act of 1824: where the old Act had given immunity for everything but intimidation and violence, the new bill—which very soon became law—restored the power of the common law over combinations, and excepted only a limited right to combine.[125] The pattern had, however, been established, and the increasing sympathy for the condition of workingmen excited by Tory reformers and early Socialists, among others, finally produced in 1871 the first Trades Union Act,[126] and in 1875, a new Combination Act[127]—both of which had the effect of legalizing all combinations of workers and masters alike, provided those combinations were formed to settle labor disputes and to negotiate hours and conditions of labor.

The law on combinations where these were labor unions or employers' associations was dominated by the statutes on the subject. On the other hand, combinations of merchants to fix prices of goods, share out a market, or otherwise limit competition, were governed by no general statutes, particularly after the laws against forestalling were repealed,[128] and therefore remained under the jurisdiction of common law. In a few instances, such combinations were indicted as criminal conspiracies at common law.[129] But after the beginning of the nineteenth century, the common law

[124] 5 Geo. IV, c. 95 (1824).

[125] 6 Geo. IV, c. 129 (1825). Cf. Dicey, op. cit. supra note 123, at 191 et seq.

[126] 34 & 35 Vic., c. 31 (1871).

[127] 38 & 39 Vic., c. 86 (1875).

[128] See pages 372–73 supra.

[129] Rex. v. Journeymen Tailors of Cambridge, 8 Mod. *10 (1721); Rex v. Eccles, 1 Leach * 274 (1783); Rex v. Mawbey, 6 T.R. *619 (1796). Cf. Winfield, The History of Conspiracy and Abuse of Legal Procedure 111–17 (1921).

came to regard an agreement between competitors to combine as analogous to a contract in restraint of trade, and judged such agreements by whether they left the parties reasonably free to act as they desired. All along, less attention was paid to whether the agreement seriously interfered with competition. In *Hearn v. Griffin*,[130] decided in 1815, the court upheld an agreement between two rival coach-owners to charge the same prices and provide no competing services; this, wrote Lord Ellenborough, "is merely a convenient mode of arranging two concerns which might otherwise ruin each other." In *Wickens v. Evans*,[131] decided in 1829, three boxmakers had agreed to divide England into areas in which each was to be the exclusive seller. This arrangement was upheld on the grounds that it was only a partial restraint of trade, one of the judges maintaining that it was "not a monopoly, except as between themselves; because every other man may come into their districts and vend his goods: all they propose is, that they shall not carry on a rivalry. . . ." In *Hilton v. Eckersly*,[132] decided in 1855, the court refused to enforce an agreement between eighteen mill owners to settle wages and hours by majority rule. The reason given was that the parties were not left free to trade on their own terms; but the case was decided before the Combination Act of 1875 authorized such agreements. In *Collins v. Locke*,[133] decided in 1879, an agreement between docking firms to distribute work and profits among themselves was upheld on the grounds that it did not unduly restrain the free action of the parties. But the modern common law on combinations in restraint of trade was established by the *Mogul Steamship* case,[134] which laid down the principle that although a trade combination might be destroyed by attack from within, it could not be successfully attacked by an outsider.

The *Mogul Steamship* case was decided two years before the *Nordenfelt* case; these two are among the chief reasons for the subsequent inability of English common law substantially to deter the growth of monopolies. The defendants in the *Mogul Steamship* case were a number of shipping lines who had formed an association, and agreed to regulate by their joint decision the number of ships each would send to Hankow or Shanghai during the brief tea-export season, the division of cargoes between those

[130] 2 Chitty 407 (1815).

[131] 3 Y. & T. 318 (1829).

[132] 6 El. & Bl. *47 (1855).

[133] 4 L.R.A.C. 674 (1855). Cf. The Shrewsbury and Birmingham Railway Co. v. The London and North-Western Railway Co., 21 L.J.Q.B. 89 (1851).

[134] Mogul Steamship Co. v. McGregor [1892] App. Cas. 25.

ships, and freight rates, to give a rebate to all shippers who dealt exclusively with members of the association, and to prohibit their agents in China from acting in the interest of competing shippers. Short of an outright merger or trust agreement, they could hardly have formed a more complete combination. The Mogul Steamship Company had been included in the association at first, were later excluded, but continued to send their ships to a Chinese port. The defendants retaliated by sending more of their ships to the port, underbidding Mogul rates, threatening to dismiss from their service agents who arranged to load Mogul ships, and circulating notices that they would not give their rebate to anyone who shipped by Mogul. The Mogul Company brought a suit for damages against the members of the association, alleging that they formed a conspiracy to injure Mogul interests. On the original action the Chief Justice, Lord Coleridge, judged for the defendants;[135] this was affirmed by the Court of Appeals, three justices dissenting;[136] and the plaintiff appealed to the House of Lords, which unanimously affirmed the decision.[137]

The principle on which all the justices agreed was that the agreement was unlawful in the sense that the courts would not enforce it, but that it was not "*contrary to law.*"[138] It would become an illegal conspiracy at common law only if it sought an unlawful end or used unlawful means; and since, as the Lord Chancellor put it, the combination had neither acted with any "malicious intention to injure rival traders," nor used any unlawful means, such as violence, intimidation, molestation, or inducing people to break contracts, it was innocent.[139] There were indeed some qualms among the justices as to whether all the means were lawful, whether the threatening notices distributed by the defendants did not amount to intimidation, but all these doubts they resolved in favor of the defendants. There were, on the other hand, no doubts at all about the propriety of defendants' ends. The defendants themselves had dismissed the public policy question, submitting that "[w]hether such combinations and agreements are on the whole beneficial or not to the public is a question not of law but of political economy as to which there will always be a difference of opinion";[140] but the court refused to accept this easy way out. Lord Bramwell maintained, on the contrary, that the public policy of free trade positively authorized such combinations: "It does seem strange," he wrote, "that to enforce freedom of trade, of action, the law

[135] 21 Q.B.D. 544 (1888).

[136] 23 Q.B.D. 598 (1889).

[137] [1892] App. Cas. 25.

[138] Ibid., at 39.

[139] Ibid., at 36–37.

[140] Ibid., at 34–35.

should punish those who make a perfectly honest agreement with a belief that it is fairly required for their protection." And he went on to suggest that combinations of employers should be given equal treatment with combinations of workmen:

> I have always said that a combination of workmen, an agreement among them to cease work except for higher wages, and a strike in consequence, was lawful at common law; perhaps not enforceable inter se, but not indictable. The Legislature has now so declared.[141]

It followed that a combination of employers was lawful at common law. The law on combinations had by this time done its full circle: earlier in the century, unions had been legalized because combinations of masters were seldom if ever punished; by the end of the century, business combinations were held to be exempt from punishment because labor unions had been legalized.

By 1890, what there had been of English common law against monopolies had become quite weak. The common law against monopoly proper had been superseded by the Statute of Monopolies. The common law against forestalling had been abolished by the statute of 1844. The common law against combinations of workmen and of masters had been overruled by the Trade Union Acts. The common law against contracts and combinations in restraint of trade alone remained in force, but it was governed by principles that condoned more than they prohibited. If monopolies were to be restrained, the common law would have to change its direction again, or legislation would have to remedy its weakness.

Legislation toward this end was provided in America, first by the antitrust laws of several states, and in 1890, by the Sherman Act. Such legislation had a firmer foundation in the United States than in England, because American common law in 1890 still contained provisions that had been struck from the English common law by statutes which had no legal effect here. Thus the common law against forestalling and engrossing was still in force in the United States; it could be made to serve purposes for which it had not been originally intended, and it was made the basis for section 2 of the Sherman Act, which prohibits "monopolizing."[142] Similarly, labor unions were not exempted from the American common law on

[141] Ibid., at 47.

[142] Adler, Monopolizing at Common Law and under Sec. II of the Sherman Antitrust Act, 31 Harv. L. Rev. 246 (1917).

combinations, and therefore the prohibition of combinations in section 1 of the Sherman Act could be used against certain activities of unions. But the Sherman Act went far beyond the common law when it authorized injured persons to sue, and the Attorney General to indict violators of the Act, making it possible to enforce competition actively. The Act was therefore much more an innovation than its authors realized. It did not, as they thought, merely declare the common law. It can almost be said to have helped create the common law, insofar as its authors' convictions helped spread the belief that the common law always expressed as much antagonism to monopoly as they wrote into the Sherman Act.

This page intentionally left blank

University of Pennsylvania
Law Review
FOUNDED 1852

Formerly
American Law Register

VOL. 115 FEBRUARY 1967 No. 4

UNCONSCIONABILITY AND THE CODE— THE EMPEROR'S NEW CLAUSE

ARTHUR ALLEN LEFF †

INTRODUCTION

This paper is devoted wholly to section 2-302 of the Uniform Commercial Code, the so-called unconscionability clause.[1] It is, however, not primarily an essay on commercial law. Rather it is intended to be a study in statutory pathology, an examination in some depth of the misdrafting of one section of a massive, codifying statute and the misinterpretations which came to surround it. The paper therefore is not intended as a commentary upon the content or drafting technique

† Assistant Professor of Law, Washington University Law School. B.A. 1956, Amherst College. LL.B. 1959, Harvard University.

[1] UNIFORM COMMERCIAL CODE § 2-302 (1962). This is the current version of the Code, which will be cited hereinafter as UCC. In the course of this paper various other versions of the Code and its predecessor statutes will be cited: 1) NATIONAL CONFERENCE OF COMMISSIONERS ON UNIFORM STATE LAWS, PROPOSED REPORT ON AND DRAFT OF A REVISED UNIFORM SALES ACT (1941) [hereinafter cited as MIMEO 1941 DRAFT]. There are two versions of the 1941 draft, one mimeographed and the other printed. Only the former contains the text of the then predecessor to § 2-302. See note 12 *infra* for further details. The printed version will be cited hereinafter as PRINTED 1941 DRAFT. 2) NATIONAL CONFERENCE OF COMMISSIONERS ON UNIFORM STATE LAWS, UNIFORM REVISED SALES ACT (Third Draft, 1943) [hereinafter cited as 1943 DRAFT]. 3) AMERICAN LAW INSTITUTE, SALES SECTIONS (SALES ACT), COUNCIL DRAFT No. 1 (1944) [hereinafter cited as FEB. 1944 DRAFT]. 4) AMERICAN LAW INSTITUTE, UNIFORM REVISED SALES ACT (Proposed Final Draft No. 1, 1944) [hereinafter cited as MAY 1944 DRAFT]. 5) AMERICAN LAW INSTITUTE, THE CODE OF COMMERCIAL LAW (1948) [hereinafter cited as 1948 DRAFT]. 6) AMERICAN LAW INSTITUTE, UNIFORM COMMERCIAL CODE (May 1949 Draft) [hereinafter cited as 1949 DRAFT]. 7) AMERICAN LAW INSTITUTE, UNIFORM COMMERCIAL CODE (Proposed Final Draft, 1950) [hereinafter cited as 1950 DRAFT]. 8) AMERICAN LAW INSTITUTE, UNIFORM COMMERCIAL CODE (Official Draft, 1952) [hereinafter cited as 1952 DRAFT].

Other drafts of the Code exist, see Braucher, *The Legislative History of the Uniform Commercial Code*, 58 COLUM. L. REV. 798 n.1 (1958), but are not important in tracing the development of the unconscionability provision of the Code.

of the Code as a whole or even of the Sales article.[2] The focus of this study is section 2-302, and excursions into other provisions of the Code are made only to help illuminate that primary target.[3] Basic to the justification of this narrow focus is the belief that such a carefully limited study will be of interest transcending that particular section's own substantive effect on the law, but that is not to say that section 2-302 was chosen at random, or that talk about its actual effect can or will be avoided here. The section was chosen because of its intrinsic interest and potential importance to both sales and contract law development,[4] already the subject of substantial controversy.[5] But the primary weight of the essay will be on section 2-302 as a thing-in-itself and how it got that way, rather than on what its operative effect might be.

Let us begin the story the way so many good stories begin, with ritual incantation: to make a contract one needs (i) parties with capacity, (ii) manifested assent, and (iii) consideration.[6] This is all very simple.[7] If these criteria are met, a party to the resulting nexus

[2] The Code has already occasioned one of the great outpourings of legal commentary in history. The following are the most extensive bibliographies of writings on the Code: BOSTON COLLEGE, UNIFORM COMMERCIAL CODE CO-ORDINATOR ANNOTATED 667-730 (1963); EZER, UNIFORM COMMERCIAL CODE BIBLIOGRAPHY—1966 (1966) ("95% of the published materials"); GOODRICH & WOLKIN, THE STORY OF THE AMERICAN LAW INSTITUTE 1923-1961 at 48-66 (1961); [1954] NEW YORK LAW REVISION COMM'N REPORT 19-46 (N.Y. Leg. Doc. No. 65(A)); WYPYSKI, THE UNIFORM COMMERCIAL CODE, A BIBLIOGRAPHY OF LEGAL ARTICLES AND PUBLICATIONS (1954).

[3] I should not like to give the impression that I consider myself the discoverer of § 2-302. I have thus far noted in excess of 130 "discussions" of the unconscionability provision in law reviews, bar journals, practice manuals, treatises and miscellaneous studies. Most of the discussions are brief and superficially descriptive, but a number have been directed primarily to § 2-302 and the unconscionability concept. See Note, 58 DICK. L. REV. 161 (1954); Note, 45 IOWA L. REV. 843 (1960); Comment, 18 U. CHI. L. REV. 146 (1950); Note, 109 U. PA. L. REV. 401 (1961); Note, 45 VA. L. REV. 583 (1959); Note, 70 YALE L.J. 453 (1961); Note, 63 YALE L.J. 560 (1954).

[4] It has been suggested that the Code's unconscionability doctrine will not be limited to the law of Sales for long, but is likely speedily to enter the general law of contracts. 5A CORBIN, CONTRACTS § 1164, at 223 (1964); Note, 65 COLUM. L. REV. 880, 891-92 (1965). See also King, *The New Conceptualism of the Uniform Commercial Code*, 10 ST. LOUIS U.L.J. 30, 39-41, 43 (1966).

[5] Noting the controversy: 1 HAWKLAND, A TRANSACTIONAL GUIDE TO THE UNIFORM COMMERCIAL CODE 44-48 (1964); Carrington, *The Uniform Commercial Code —Sales, Bulk Sales and Documents of Title*, 15 WYO. L.J. 1, 7 (1961); Hawkland, *In re Articles 1, 2 and 6*, 28 TEMPLE L.Q. 512, 513 (1955); Kripke, *The Principles Underlying the Drafting of the Uniform Commercial Code*, 1962 U. ILL. L.F. 321, 324. Critical or at least questioning: Buerger, *The Sales Article of the Proposed Uniform Commercial Code*, 23 N.Y.S.B. BULL. 116, 120 (1951); Douglass, *Discussion on Sales as Proposed in the Uniform Commercial Code*, 21 OKLA. B.A.J. 808, 810 (1950); Goodwin, *How the Adoption of the Uniform Commercial Code Would Affect the Law of Sales in Oregon*, 30 ORE. L. REV. 212, 213-14 (1951); Ireton, *The Commercial Code*, 22 MISS. L.J. 273, 280 (1950); King, *Suggested Changes in the Uniform Commercial Code—Sales*, 33 ORE. L. REV. 113, 115-16 (1954); Levy, *A Study of the Uniform Commercial Code—Sales*, 58 COM. L.J. 329, 331 (1953).

[6] RESTATEMENT, CONTRACTS § 19 (1932).

[7] This simplicity is, of course, of a rather special kind. Robert Frost once remarked (at a "saying" of his poetry): "e equals mc²; what's so hard about that? Of course, what e, m and c are is harder.

who has made promises is obligated to carry them out, unless he can maintain successfully one of the standard contract-law defenses, such as fraud, duress, mistake, impossibility or illegality.[8] These "defenses" might be classified in divers ways to serve various analytical purposes. For our particular needs, however, there is a simple way of grouping them which is signally illuminating: some of these defenses have to do with the *process of contracting* and others have to do with the resulting *contract*. When fraud and duress are involved, for instance, the focus of attention is on what took place between the parties at the making of the contract. With illegality, on the other hand, the material question is instead the content of the contract once "made."[9] The law may legitimately be interested both in the way agreements come about and in what they provide. A "contract" gotten at gunpoint may be avoided; a classic dicker over Dobbin may come to naught if horse owning is illegal. Hereafter, to distinguish the two interests, I shall often refer to bargaining naughtiness as "procedural unconscionability," and to evils in the resulting contract as "substantive unconscionability."

Getting down to cases, section 2-302 of the Uniform Commercial Code provides in its entirety as follows:

> Section 2-302. Unconscionable Contract or Clause.
>
> (1) If the court as a matter of law finds the contract or any clause of the contract to have been unconscionable at the time it was made the court may refuse to enforce the contract, or it may enforce the remainder of the contract without the unconscionable clause, or it may so limit the application of any unconscionable clause as to avoid any unconscionable result.
>
> (2) When it is claimed or appears to the court that the contract or any clause thereof may be unconscionable the parties shall be afforded a reasonable opportunity to present evidence as to its commercial setting, purpose, and effect to aid the court in making the determination.

If reading this section makes anything clear it is that reading this section alone makes nothing clear about the meaning of "unconscionable" except perhaps that it is pejorative.[10] More particularly, one cannot

[8] See RESTATEMENT, CONTRACTS §§ 19(d), 454-609 (1932).

[9] It is possible in some cases for the contracting process to be illegal while performance of the contract is not. See 6A CORBIN, CONTRACTS § 1373 (1962).

[10] As one would suspect from its linguistic structure alone, which is the negativing of the root concept of "conscience." See WEBSTER, NEW INTERNATIONAL DICTIONARY OF THE ENGLISH LANGUAGE 2763 (2d ed. 1957). The examples of its use collected in 11 OXFORD ENGLISH DICTIONARY U-99 (1933) show the word to have been used through the ages as a rather generalized pejorative intensifier, a wide-gauge "snarl word." See HAYAKAWA, LANGUAGE IN ACTION 76-79 (1941).

tell from the statute whether the key concept is something to be *predicated on* the bargaining process or on the bargain or on some combination of the two, that is, to use our terminology, whether it is procedural or substantive. Nonetheless, determining whether the section's target is a species of quasi-fraud or quasi-duress, or whether it is a species of quasi-illegality, is obviously the key to the bite and scope of the provision.

One central thesis of this essay is that the draftsmen [11] failed fully to appreciate the significance of the unconscionability concept's necessary procedure-substance dichotomy and that such failure is one of the primary reasons for section 2-302's final amorphous unintelligibility and its accompanying commentary's final irrelevance. This I think can most clearly be shown by an examination in detail of the drafting history of the provision and its accompanying comments, from the beginning (prior to 1941) to the present version. The examination will proceed first from the point of view of what that history discloses about the transformations of procedural unconscionability, and then the focus will shift to substantive unconscionability. Thereafter, I shall examine the equity-specific performance "unconscionability" doctrine, to show its total inapplicability to the problems dealt with in the Code, hence pointing out the irrelevance of substantially all of the standard commentary on the section. I shall close by examining the reported cases thus far affected by section 2-302, and their dangerous

[11] I shall use "draftsmen" throughout to refer to that imaginary construct which corporately produced the final Code and the final version of § 2-302. From time to time I shall use the singular form "draftsman," to refer to the late Karl Llewellyn who, at least at the earliest drafting stages, did the major share of the actual drafting, especially of the Sales article. See Braucher, *supra* note 1, at 800; Mooney, *Old Kontract Principles and Karl's New Kode: An Essay on the Jurisprudence of Our New Commercial Law*, 11 VILL. L. REV. 213, 223 (1966). And see the early and amusing evaluation in Franklin, *On the Legal Method of the Uniform Commercial Code*, 16 LAW & CONTEMP. PROB. 330 (1951), where the Code is continually referred to as the "lex Llewellyn." *But see* Llewellyn, *Why A Commercial Code?*, 22 TENN. L. REV. 779, 784 (1953): "there are so many beautiful ideas I tried to get in . . . but I was voted down."

It became an article of faith for the defenders of the Code to assert that no single man or group had a monopoly of the drafting of the Code, especially (during the height of the adoption push) not law professors. There are a goodly number of articles on who "really" drafted the Code, taking somewhat divergent views. See, suggesting that the professors really ruled, Beutel, *The Proposed Uniform (?) Commercial Code Should Not Be Adopted in Ohio*, 14 OHIO ST. L.J. 3, 6-7 (1953) (conspiracy between professors and *successful* lawyers); Kripke, *The Principles Underlying the Drafting of the Uniform Commercial Code*, 1962 U. ILL. L.F. 321, 321-28; Levy, *A Study of the Uniform Commercial Code—Sales*, 58 COM. L.J. 329 (1953). Among the works defending against this academic-orientation canard are Godfrey, *Preview of the Uniform Commercial Code*, 16 ALBANY L. REV. 22, 25-26 (1952); Kuhns, *Uniform Commercial Code*, 16 TEX. B.J. 67, 68 (1953); Malcolm, *The Uniform Commercial Code*, 39 ORE. L. REV. 318, 323 (1960); Mentschikoff, *The Uniform Commercial Code, An Experiment in Democracy in Drafting*, 36 A.B.A.J. 419 (1950); Note, 65 COLUM. L. REV. 880, 887 (1965).

The part Professor Llewellyn played in the final form of § 2-302 is hard to assess. In his later writings about it he expressed neither hostility to the section nor much faith in it. See LLEWELLYN, THE COMMON LAW TRADITION 370-71 (1960).

(though understandable) tendency. The central purpose of the paper will be to illustrate the progressive abstraction, attentuation and eventual destruction of meaning in an important single statutory provision, in response to pressures the nature of which can only be guessed.

PROCEDURAL UNCONSCIONABILITY

Drafting History

In 1941 there appeared publicly for the first time the provision which eventually became section 2-302 of the Code.[12] In that earliest draft at least, there was substantial evidence that the draftsman intended to provide that if a contract or portion thereof were in fact the subject of some (not quite specified) level of particularized bargaining, it would be safe from judicial rewriting. This original tack must be emphasized. From the beginning the procedural unconscionability question was not posed in terms of what bargaining conduct, if any, would vitiate the agreement, but rather in terms of whether there was bargaining conduct sufficient to *insulate* from judicial interference a contract which was, arguably, substantively "unconscionable." The draft provision at one point indicated that view quite distinctly:

> When both of the parties have so directed their attention to a particular point that . . . variance from this Act may fairly be regarded as the deliberate desire of both, and as reflecting a considered bargain on that particular point . . . the legislature recognizes that policy in general requires the parties' particular bargain to control.[13]

Other portions of the provision helped to reinforce this idea that bargaining of some dimension would validate any contractual term, for instance the clear statement that the section's policy was "to aid

[12] MIMEO 1941 DRAFT § 1-C. This draft is identified in a covering letter dated September 5, 1941, from Professor Llewellyn (then at Columbia Law School) to Professor Underhill Moore at Yale Law School, as a "Second Draft of a Revised Sales Act, for the Committee's discussion . . . Sept. 19-22." Section 1-C is new in this draft. (The letter is bound in with the Yale Law School Library's copy of the mimeographed 1941 draft.)

The 1941 mimeographed version is apparently not very widely available. For our purposes the mimeographed 1941 draft is particularly important, for in the printed 1941 version (copies of which exist in abundance), § 1-C is omitted, and at the point at which the section would have appeared, there appears: "[Section 1-C. (New to Sales Act). *Form Contracts and Particularized Terms.* This section was withdrawn by the Committee . . . .]"

[13] MIMEO 1941 DRAFT § 1-C(1)(b). Section 1-C was no ordinary statutory provision. It runs ninety-nine lines (almost three full pages), is accompanied by almost five pages of commentary and the "Report" accompanying the draft and its comment devotes the greater part of an additional four pages primarily to its explication.

The draftsman's defense against any suggestion that such fullness might approach fulsomeness may be found as part of § 1-C, comment B(1). It is not reprinted in PRINTED 1941 DRAFT.

and foster any considered and deliberate action of the parties." [14] Moreover, the accompanying comments further reinforced that interpretation:

> *The principle of freedom of bargain* is a principle of freedom of *intended* bargain. It requires what the parties' [*sic*] have bargained out to stand as the parties have shaped it, subject only to certain overriding rules of public policy. . . . Displacement of these balanced backgrounds [provided by the Code] is not to be assumed as intended unless deliberate intent is shown that they shall be displaced . . . .[15]

On the other hand, there were hints that perhaps there were some contracts or clauses which, under the general rubric of "unconscionability," would not be enforced regardless of what the bargaining process was like. For instance, the sentence quoted above about the section's policy to "aid and foster the considered and deliberate action of the parties" closed with what might have been a limiting condition: "in substituting for the general rules of this Act a fair and balanced set of provisions more particularly fitted to the needs of any particular trade or situation." [16] This might mean that "considered and deliberate action" which resulted in unfair and unbalanced provisions not required by the "needs" of any particular trade or business would not be binding on the parties. But how that situation was to be treated under the section was not clear, even to the draftsman, one suspects. The idea might have been that the unbalanced nature of the resulting terms was evidence that, despite appearances, there was *not* after all the requisite bargaining. At another point the section provided:

> If the bloc [of form provisions] as a whole is shown affirmatively to work a displacement or modification of the provisions of this Act in an unfair and unbalanced fashion not required by the circumstances of the trade, then the party claiming application of any particular provision in

[14] MIMEO 1941 DRAFT § 1-C(1)(e). In addition to this picture of what man-to-man bargaining would be sufficient to validate a departure from the Code rules, the 1941 Draft also envisioned valid form contracts arrived at through groups bargaining for a particular trade or industry. If a form contract were thus arrived at to govern a particular trade, and *that* bargaining procedure were "fair," then the modified contract would be impregnable.

The draft provided:

> The legislature also recognizes that particular trades and situations often require extensive special regulation in a manner departing from the general provisions of this Act, and that speed and convenience in transacting business may require such extensive departures to be incorporated into a general form contract, or into "rules" to which particular transactions are made subject, although the details of such "rules" or forms are not so deliberated on and bargained by the two parties when they are closing an individual transaction as to become particularized terms of the bargain.

MIMEO 1941 DRAFT § 1-C(1)(c).

This validation-by-proxy technique totally disappeared after the 1941 draft.

[15] MIMEO 1941 DRAFT § 1-C, comment A(3), at 18. (Emphasis in original.)

[16] MIMEO 1941 DRAFT § 1-C(1)(e).

> such bloc must show that the other party, with due knowledge of the contents of that particular provision, intended that provision to displace or modify the relevant provision of this Act in regard to the particular transaction.[17]

In fact, the section and its accompanying commentary really spoke throughout as if it were inconceivable that there could exist simultaneously both particularized bargaining and an unfair contract. The idea seems almost to have been that if a clause with which businessmen have been living looks unfair to an outsider it is only because he fails to understand the particular context.

> Many groups of clauses in very frequent use in the Sales field are utterly one-sided, but are, for all that, entirely fair because they correct in a reasonable way an unfortunate condition in the law.[18]

Thus, it is just not clear under the 1941 version what result would be reached with respect to a provision, or a block of provisions, or an entire contract which not only looked one-sided but *was* one-sided. In other words, if the complaining party, at the time he entered into the contract, knew the nature of the "intended bargain," [19] and had "due knowledge of the contents of that particular provision," [20] but was in no bargaining position to get any changes made, would a contract entered into under such circumstances, no matter what its terms, be safe?

The only fair answer, I think, is that one cannot tell for sure. No doubt the overall drift of the section was that contracts ought to be "fair and balanced" no matter how the parties bargained, but at least as far as its explicit language went, the draft section also was committed to the view that explicit bargaining would insulate a contract: if A and B actually bargained over each clause of a contract, and each came out the way A wanted it, the contract would stand even if extremely onerous to B. One of the central problems of the section, therefore, arose as early as 1941: what, *if anything*, will insulate a contract from 2-302? The 1941 draft used several locutions to envisage an apparently very stringent bargaining standard which might succeed, but the problem remained radically unsolved.

The 1941 draft of the section did not survive its first exposure to the light. It was "withdrawn by the [drafting] committee" because the "machinery for administration thus far developed" was thought to be "inadequate" and "too unreckonable to be in keeping with the lines

[17] MIMEO 1941 DRAFT § 1-C(2)(a)(i).
[18] MIMEO 1941 DRAFT § 1-C, comment A(6), at 19.
[19] MIMEO 1941 DRAFT § 1-C, comment A(3), at 18.
[20] MIMEO 1941 DRAFT § 1-C(2)(a)(i).

of the draft." [21] Perhaps because of these "machinery" and "reckonability" problems, when the third draft of the Uniform Revised Sales Act came out in August 1943, matters on the unconscionability front were materially changed. This new predecessor to 2-302 read as follows:

> SECTION 24. FORM CLAUSES, CONSCIONABLE AND UNCONSCIONABLE. (1) A party who signs or accepts a writing evidencing a contract for sale which contains or incorporates one or more form clauses presented by the other party is bound by them unless the writing when read in its entirety including the form clauses is an unconscionable contract and he has not in fact read the form clauses before contracting, except that a merchant who signs and returns such a writing after having had a reasonable time to read it is bound by it.[22]

This version clearly lacked much of the *salvator mundi* touch of the 1941 draft. The section was explicitly made applicable to "form clauses" only.[23] The power to bind someone to something which he had merely read, seemingly extinguished in the 1941 draft, was here clearly resuscitated. Moreover, for merchants,[24] under the 1943 version, not even reading was necessary; so long as they had had sufficient opportunity to read, they were bound. The accompanying comment [25] was quite explicit. Where the 1941 comment spoke in terms of "deliberate intent," [26] the 1943 comment began:

> The situation which gives rise to the section is the increasing use of forms prepared by one party which are not in fact *examined* by the other at the time of contracting . . . .[27]

In short, between 1941 and 1943 the provision moved from a search for words to paint pictures of haggling to a search for words for expressing merely looking and reading.[28]

---

[21] PRINTED 1941 DRAFT 51-52.

[22] 1943 DRAFT § 24. This version contained two additional subsections, subsection (2) providing that any form recital that clauses were read was to be disregarded, and subsection (3) detailing the alternative procedures open to the court once it made a finding of unconscionability (including reformation, excision and nullification).

[23] The heading of MIMEO 1941 DRAFT § 1-C read: "Declaration of Policy, and Procedure with Regard to Displacement of Single Provisions or Groups of Provisions by Agreement." The heading of 1943 DRAFT read: "Form Clauses, Conscionable and Unconscionable."

[24] See UCC § 2-104(1) for the Code's definition of "merchant."

[25] Strictly speaking, there is no "comment" to this 1943 draft. There is, however, a mimeographed document entitled *Informal Appendix to Revised Uniform Sales Act, Third Draft, 1943, Tentative Sketch of Material for Comments* (1943). This appendix was "submitted on the sole responsibility of the draftsman," *i.e.*, Professor Llewellyn.

[26] MIMEO 1941 DRAFT § 1-C, comment A(3), at 19.

[27] *Informal Appendix, supra* note 25, at 11. (Emphasis added.)

[28] Even here, however, one must beware. The end of that portion of the "Informal Appendix" devoted to § 24 (the then foetal form of § 2-302) referred to "matters which were not particularly *discussed* by the parties." *Id.* at 12. (Emphasis added.) That seems to imply that the draftsman may still have had in mind something more than mere reading.

This 1943 draft, however, was destined to be the section's high point of permissiveness. By February 1944, when the next published draft was submitted to the Council of the American Law Institute, while the special merchant's provision remained intact, the clause "and he has not in fact read the form clauses before contracting" was eliminated.[29] By May 1944, when the draft was submitted by the Council to the Institute's membership (by which time the draftsmen had hopefully labelled it "Proposed Final Draft No. 1"), even the special merchant provision was gone.[30]

Oddly enough, it is not immediately clear what the effect of those changes might have been. The reference to "reading" in the prior draft might have been eliminated because the draftsmen felt that mere reading ought not to be a sufficient insulating factor, but that something further ought also be shown to save the contract—perhaps that the clause not only be read but understood,[31] or perhaps even that actual haggling have taken place. That is, the elimination of the reference to "reading" may have been an attempt merely to return the section to the stringent procedural requirements of the 1941 draft.

On the other hand, the elimination of the reference to reading might be understood more simply. When one draft of a statute provides that a clause will be vulnerable if it is *both* (1) "unconscionable" *and* (2) not read, and the next draft removes any reference to reading, one might fairly conclude that when a contract or clause is unconscionable, it is unconscionable, and *no* amount of reading or bargaining or understanding will make any difference. Such a reading leads, obviously, to the more radical of the positions which may have been contained in the 1941 draft, that even contracts or clauses which were expressly bargained about might be stricken or modified if "unconscionable." Admittedly this seems not to have been the major thrust of the earliest draft. Under the 1941 formulation it appeared that *some* amount of bargaining fullness would rescue *any* contract or clause.[32] But the May 1944 draft might have meant that substantive unconscionability alone could vitiate any provision, and no amount of procedural "superconscionability" could save it.

This interpretation, however, runs into an objection other than radicalism. If the bargaining procedure were to be considered irrele-

[29] See FEB. 1944 DRAFT § 23.

[30] See MAY 1944 DRAFT § 23.

[31] *Cf.* MIMEO 1941 DRAFT § 1-C(2)(a)(i) with its suggestive "due knowledge of the contents of that particular provision."

[32] Subject, of course, to the usual illegality limitations. *Cf.* MIMEO 1941 DRAFT § 1-C, comment A(3), at 18:

> *The principle of freedom of bargain* . . . . requires what the parties' [*sic*] have bargained out to stand as the parties have shaped it, subject only to certain overriding rules of public policy. (Emphasis in original.)

vant to the conscionability determination, why would the section be explicitly limited to "form clauses"?[33] Should a bargained-about form clause be any worse off under the section than a bargained-about clause created for the occasion? It seems somehow an unlikely result that if a party arrives with a blank sheet of paper and writes clause X upon it, clause X will be invulnerable no matter what it says, but if he comes with a form on which appears clause X, even if he and the other party dicker over it specifically, clause X may be later stricken by the court. The fact that a particular clause is part of a form has no bearing upon its effect, that is, upon its substantive conscionability; its form-genesis is relevant *only* if the nature of the bargaining process is relevant to the section.

The 1944 draft, therefore, was highly unstable. The reference to "reading" from the prior draft was gone; no new standard for sufficient bargaining was supplied; and yet the section was applicable by its terms only to form contracts, thus making procedural factors relevant. This obvious tension between per se unconscionability and unconscionability to which the bargaining process was material could not remain long unsettled. In fact, the most surprising thing is that the draftsmen managed to put off the choice as long as they did, until 1948. In that year's version of the unconscionability section the ambiguity contained in the previous drafts was dispelled. The section read in its entirety as follows:

> Section 2-302. Unconscionable Contract or Clause.
>
> (1) If the court finds the contract to be unconscionable, it may refuse to enforce the contract or strike any unconscionable clauses and enforce the rest of the contract or substitute for the stricken clause such provisions as would be implied under this Act if the stricken clause had never existed.
>
> (2) A contract not unconscionable in its entirety but containing an unconscionable clause, whether a form clause or not, may be enforced with any such clause stricken.[34]

There is one hint in the section (without reference to its accompanying comment) which indicates that a doctrine of per se unconscionability had been chosen: the limited application of the section only to "form"

---

[33] See note 23 *supra* and accompanying text.

[34] 1948 DRAFT § 2-302. The changes in this version may have been conceived well before 1948. There is a mimeographed version of the sales article extant dated "as of 4/1/46" which is identical with the 1948 and 1949 versions (though possessing no comments). See AMERICAN LAW INSTITUTE, UNIFORM REVISED SALES ACT (1946).

It may be well to note here that there likely are other vagrant mimeographed versions of the Code or portions thereof which I have not seen, my searching having been limited substantially to the Columbia, New York University, Yale and Washington University Law Libraries.

contracts was eliminated.[35] In fact, this draft's sole reference to form clauses was inserted only to make clear that whether the clause was a form clause or not was irrelevant.[36] Thus, though once again the language of the statute falls somewhat short of limpidity, one gains the impression that the draftsmen might have decided to make "unconscionability" in at least some cases independent of the bargaining process.

And for once the natural inference drawn from the section was clinched by the accompanying commentary. The third comment to this 1949 version dealt specifically with procedural unconscionability. It read as follows:

> A common type of unconscionable clause within this section consists of cases in which the contents of the questionable clause were never actually discussed or bargained out by the parties and as a result the clause was included in the agreement without one party's attention ever having been directed specifically to it. This situation arises most frequently with respect to "form" contracts where the attention of the other party is addressed to the bargained terms which are filled in.[37]

What is most noteworthy about this comment, I think, especially in the light of what the fourth comment was to say, is that the draftsmen apparently found it exceedingly difficult to pin down exactly what was worrying them concerning procedural unconscionability. In one short paragraph they described the bargaining vice as a failure of discussion, a failure of bargaining and a failure to have one's attention "directed specifically" to a clause. Perhaps it was this difficulty in pinning down the naughty bargaining conduct which prompted the fourth comment, which read as follows:

> Another common type of situation arising in connection with unconscionable contracts or clauses consists of cases where one party has deliberately entered into a lopsided bargain with full knowledge and awareness and has actually assented to clauses which are unconscionable in effect against him. In such cases this Article goes on the theory that sales contracts have as their legally necessary effect certain minimum incidents set forth in this Article despite any agreement

[35] This version is the first printed version which does not mention "form" contracts in the title, but instead substitutes a reference to "unconscionable" contracts. Permit me to suggest that you close your eyes and try to *picture*, respectively, a "form" contract and then an "unconscionable" contract.

[36] 1948 DRAFT § 23. This, it should be pointed out, is in itself a pretty peculiar form of drafting, to negative an impression which could have been gained only by having read previous drafts of a section.

[37] 1949 DRAFT § 2-302, comment 3. The text of the section in the 1949 DRAFT (which contained comments) was identical with 1948 DRAFT § 23.

> of the parties to the contrary. The question primarily is whether or not a contract for sale in a business sense was intended. If so, then the transaction is governed by this Article and its minimum legal effects are laid down by the law as embodied in this Article. Therefore, the court may, under this section, refuse to enforce the clause or agreement as unconscionable and declare that the provisions of this Article be made operative instead.[88]

This comment seems to have the effect (even if not conscious) of making the immediately previous one irrelevant and atavistic. Interpreting section 2-302 along the lines suggested by this fourth comment has the natural effect of solving, at one swoop, substantially all of the problems one might have in deciding upon, classifying and conveying in language to another just what contracting *conduct* is objectionable. One merely drops the question. The process of getting the clause becomes unimportant. An unconscionable clause is unconscionable. But the price paid for this facile solution is to increase immensely the weight which has to be borne by the definition of "unconscionable" as a substantive thing, because now that decision may be made without reference to the bargaining process at all. That does not mean that it must be so made, but the net result is to make it possible under the section to strike a single provision in a contract even if it had been specifically bargained about and even if it were not forbidden by any established doctrine of illegality or public policy, solely on the basis of an *ad hoc* judicial determination of substantive "unconscionability."

Such a position, even though fraught with difficulties and somewhat radical, had all the virtues of clarity. Moreover, the world would not have come to an end if that position on the definition of "unconscionability" had been the one finally adopted. As a matter of statutory draftsmanship, after wrestling with the serious policy problems involved, one might have decided to make the bargaining process irrelevant. The draftsmen in the earlier drafts had tried to make the bargaining process relevant, and had encountered immense difficulties in describing the mechanics and details of that relevance. This 1949 draft cut the developing Gordian knot by saying, in effect, that an unconscionable clause is an unconscionable clause, no matter how it got into the contract. The policy determination was made, in effect, that one could use his superior bargaining power only so far. A legislature being presented with the 1949 draft would have had a fighting chance of knowing what it was being called upon to import into the law of Sales.

Alas, the draftsmen's impulse toward transparency of intention was but ephemeral. It lasted only until the next printed draft of the

[88] 1949 DRAFT § 2-302, comment 4.

Code came out in May 1950. The statute itself remained almost unchanged from its 1949 incarnation.[39] But the comments, oh my, the comments. As a starter, comments 3 and 4 from the prior version, the two comments which explicitly discussed and distinguished substantive and procedural unconscionability, were totally deleted. There was substituted, however, a newly minted first comment, which read in its entirety as follows:

> This section is intended to make it possible for the courts to police explicitly against the contracts or clauses which they find to be unconscionable. In the past such policing has been accomplished by adverse construction of language, by manipulation of the rules of offer and acceptance or by determination that the clause is contrary to public policy or to the dominant purpose of the contract. This section is intended to allow the court to pass on the unconscionability of the contract or particular clause therein and to make a conclusion of law as to its unconscionability. The basic test is whether in the light of the general commercial background and the commercial needs of the particular trade or case the clauses involved are so one-sided as not to be expected to be included in the agreement. The principal [*sic*] is one of prevention of unfair surprises and not of disturbance of allocation of risks because of superior bargaining power. The underlying basis of this section is illustrated by the results in cases such as the following . . . .[40]

As I shall discuss anon, this new comment raised substantial problems through its obfuscation of what "unconscionability" as a substantive thing applicable to a single contractual provision might be. But equally significant was the diffusion of the section's attitude toward contracting conduct. Briefly put, is the manner in which a provision gets into a contract relevant or not? If the contracting process is relevant, what standards does one use to judge the adequacy of that process? Is "reading" enough (can you or can't you be surprised by what you have read?) or "understanding" or "bargaining"? If some form of bargaining over a specific clause goes on, but the seller can and does adopt a take-it-or-leave-it position, is the buyer bound if he takes it? The important thing is not so much that the comment to the 1950 version does not clearly answer those particular questions, but that it clearly replaced a draft which did.

---

[39] The second subsection of the 1949 DRAFT § 2-302, see text at note 34 *supra*, was eliminated, the first subsection having added to it the language italicized below:

"If the court finds the contract *or any clause of the contract* to be unconscionable, it may . . . ." This was tighter drafting, certainly, but made no change of substance.

[40] 1950 DRAFT § 2-302, comment 1. Following the colon were the ten cases cited in the current version of the Code.

The 1949 version's confrontation of the difficulty was to be the last; succeeding drafts of 2-302 were to back further and further away from any stand on the relevance of contracting procedure to a finding of "unconscionability." While the section itself did not change in any manner material to that problem after the 1950 changes, the comments did—subtly perhaps, but importantly.[41] For instance, the 1950 comment had described an unconscionable clause as one "so one sided as not to be expected." The comment to the 1952 draft, however, condemned instead clauses "so one sided as to be unconscionable." [42] This particular transformation I find most instructive on the development of 2-302's language in general. The 1950 comment had pointed to a recognizable human situation; it had, if you will, a dramatic situation somewhere behind it. It may have been impossible to tell in advance what clause might turn out to be so unexpectable as to be *unfairly* surprising, but at least it was clear that one was looking for one of the indicia of surprise—a dropped jaw, perhaps. Some variation from what a contracting party might reasonably have been lulled into expecting (or, more likely, not expecting) was the focus. That would be a scene describing the interaction of real people. Obviously relevant to unconscionability posed as a question of "surprise" would be whether the clause ought to have been pointed out especially, or explained, or at the very least not hidden in fine print and verbal complexity. The test might have been stated, "if he had read this clause, and if he had understood it, what is it likely that he would have done?" If the answer were "exclaimed" or "questioned" or even perhaps "looked quizzical" (the expected reaction need not have been at the level of a silent-movie seduction) then there might arguably have been enough wrong with such a clause's method of importation into the contract to justify its lancing. The 1950 comment at least made the question one of a person's state of mind, and its factual justification. But when it was decided in the 1952 draft to describe unconscionability as "so one-

[41] This propensity of the draftsmen to make material changes in the Code by modifying the comment rather than the statute has not gone unnoticed. See Surrency, *Research in the Uniform Commercial Code*, 1962 U. ILL. L.F. 404, 408; *Report on Article 2—Sales by Certain Members of the Faculty of the Harvard Law School* [Professors Braucher, Kaplan, McCurdy & Sutherland], 6 BUS. LAW. 151, 153 (1951); Note, 71 HARV. L. REV. 674, 686 (1958).

1950 DRAFT provided that the comments might "be consulted by the courts to determine the underlying reasons, purposes and policies of this Act and may be used as a guide in its construction and operation." 1950 DRAFT § 1-102(2). The next draft provided that the comments might be consulted "in the construction and application" of the Code, "but if text and comment conflict text controls." 1952 DRAFT § 1-102(3)(f). The present version of the Code has no provision dealing with the status of the comments at all. Surrency, *supra* at 407-08, suggests that it was omitted after 1952 "because the old comments were out of date" and the draftsmen didn't know when they would be able to produce new ones. The present commentary to § 2-302 is substantially the same as it was in 1950.

[42] *Compare* 1950 DRAFT § 2-302, comment 1, *with* 1952 DRAFT § 2-302, comment 1.

sided as to be unconscionable," all dramatic focus was destroyed. The movement of the drafting was from definition in terms of drama to definition in terms of abstraction. By 1952 unconscionability was defined in terms of itself.

Still another major change was made in the 1952 comment. To the draft as it appeared in 1950 the material indicated by italics below was added:

> The principle is one of the prevention of *oppression and* unfair surprise (*Cf. Campbell Soup Co. v. Wentz, 172 F. 2d 80, 3d Cir. 1948*) and not of disturbance of allocation of risks because of superior bargaining power.[43]

The relevance, if any, of the *Campbell Soup* decision and the doctrine of equity unconscionability will be discussed shortly. At this point, however, while the historical progression of the treatment of procedural unconscionability is being surveyed, it is especially illuminating to discuss what "oppression" might possibly have meant. Given the emerging diffusing trend of the statute brought to its peak in this 1952 draft,[44] it should come as no surprise to anyone to discover that the word "oppression," apparently chosen to clarify the meaning of unconscionability, should be almost perfectly ambiguous. Oppression, strictly as a linguistic and syntactical matter, might refer to what took place between the parties at the time they entered into the contract in question (a sort of quasi-duress), or it might just as well refer to the effect of that contract upon the complaining party. As it happens, it is not easy to think of a word better designed to leave in a state of perfect uncertainty whether the focus of the section was to be upon the contracting process or the contract.

If one takes the position that "oppression" refers to the nature of the contract rather than to the contracting process, then the word may add to one's feeling for what "unconscionability" might be: it is something that is not only unexpected but hard on the complaining party. That harshness should be a component of unconscionability will hardly come as startling illumination to anyone, but it does add some explicit coloration to an implicit expectation. If, however, "oppression" describes something in the bargaining process, one is merely more puzzled. Prior to its appearance one would have, under the guidance of the reference to "unfair surprises," focused his attention upon various modes of deception which might have been practiced on the

[43] 1952 DRAFT § 2-302, comment 1.

[44] Note, for instance, this subtle linguistic modification: the 1950 draft's "surprises" became in the 1952 draft "surprise." This is a nice example of the progressive regression of § 2-302's language from recognizable commercial "plot" to abstraction. Substituting "surprise" for "surprises" has much the same effect as substituting the abstract plural "man" for the pictorial plural "men."

complaining party. One would have looked to factors such as absence of opportunity to read or ability to read, the size of the type used, the unnecessary verbal complexity of the provision in question and so forth. What factors would suffice to do the trick might not be clear, but the relevant inquiry would have been intelligibly circumscribed.[45] With the suggestion that "oppression" was to be henceforth relevant, however, there appeared a new dimension. It was as if the comment had said that if for some reason the aggrieved party could not effectively have objected to the provision in question, even if he knew about it and understood it, that is, *even if he were not surprised*, then the provision would still be destructible as unconscionable. What could be clearer?

Well, what could be clearer, and what in fact was clearer, was the statement in the 1949 draft comment that it was the intention of the draftsmen to cover those clauses which in fact were totally bargained but just too harsh to permit.[46] That particularly explicit comment, however, was eliminated very shortly after it appeared. Is one to take that the gist of that comment was deemed to have returned with all of its vigor in this new compressed form? I am easily churlish enough to suggest that drafting compression has its limits, and that if one were trying to convey such a signally radical position it would have been well to do so in a somewhat less Delphic manner than by the unexplained insertion of the single word "oppression." Moreover, even this circuitous implication that the full meaning of "oppression" encompassed "forced by strong bargaining" is somewhat lessened by the presence in the 1952 draft of an element which did not appear in the 1949 version, the express disclaimer of any intention to meddle with "superior bargaining power."[47] Since "true" duress expectedly remained an available defense in commercial contracts even after the adoption of the Code,[48] "oppression" must lie somewhere between duress and superior bargaining power, a rather narrow niche indeed.[49] Why all this ambiguity?

---

[45] One would emphasize in any opinion the factors which would prevent the complaining party from reading and understanding, for instance, general mechanical reading difficulties, Siegelman v. Cunard White Star Ltd., 221 F.2d 189, 208-10 (2d Cir. 1955) (appendix showing back of steamship ticket), or personal reading difficulties, Fricke v. Isbrandtson Co., 151 F. Supp. 465, 468 (S.D.N.Y. 1957) (steamship ticket in language party didn't know). This focus is behind the baggage-check cases, see, *e.g.*, Klar v. H. & M. Parcel Room, Inc., 270 App. Div. 538, 61 N.Y.S.2d 285 (1946), *aff'd mem.*, 296 N.Y. 1044, 73 N.E.2d 912 (1947), and the treatment often accorded generally to fine print. See Vogel & Bernstein, *Fine Print*, 21 BUS. LAW. 544 (1966); Note, 63 HARV. L. REV. 494 (1950).

[46] 1949 DRAFT § 2-302, comment 4.

[47] 1952 DRAFT § 2-302, comment 1.

[48] See UCC § 1-103.

[49] If it is intended to be a shorthand way of referring to some concept like "business duress" or "duress of goods," it is an almost grotesquely foreshortened way of

The answer, I think, is reasonably clear. The draftsmen were faced with several possibilities. They could have said that if a certain level of bargaining elaborateness were reached, *any* resulting contract (short of illegality) would be invulnerable to later judicial meddling. That, however, would most likely have necessitated some fuller description of what type of bargaining procedure was envisioned as sufficiently immunizing. That is, they would have had to return to what seems to have been the basic conception (though not necessarily to the exact language or to the discursive style) of the original 1941 version. This, as the earliest draft itself showed, presented exceedingly difficult drafting problems. Alternatively, the draftsmen could finally have espoused the position taken in the 1949 draft, that there were some contractual provisions, presently unspecifiable, which could not be permitted under the Code no matter how fully bargained between the parties. This position, however, might well have been unacceptable to important backers of the Code [50] (not to mention to legislatures) if it had been set forth in the high relief in which it was graven in the 1949 comment. Thus faced with a dilemma, the difficulty of the first alternative and the unpopularity of the second, the draftsmen opted for a third solution. They fudged.

The Official-Comment Cases

There are clues, however, that illumine the draftmen's actual conception of procedural unconscionability, public equivocating notwith-

vouching in those vexed and complex topics. On the niceties of commercial "duress," see WOODWARD, THE LAW OF QUASI-CONTRACTS §§ 211-18 (1913); Dalzell, *Duress by Economic Pressure*, 20 N.C.L. REV. 237, 341 (1942); Dawson, *Economic Duress and the Fair Exchange in French and German Law*, 11 TUL. L. REV. 345, 12 *id.* 42 (1937); Dawson, *Economic Duress—An Essay in Perspective*, 45 MICH. L. REV. 253 (1947); Annot., 79 A.L.R. 655 (1932); Note, 15 N.C.L. REV. 412 (1937). *Cf.* Note, 20 COLUM. L. REV. 80 (1920) (threat of litigation).

Somewhat peripheral for our purposes, but of great interest on the general problem of duress and quasi-duress in a contract context, are Hale's two articles, Hale, *Bargaining, Duress, and Economic Liberty*, 43 COLUM. L. REV. 603 (1943); Hale, *Coercion and Distribution in a Supposedly Non-Coercive State*, 38 POL. SCI. Q. 470 (1923), with their careful discussion of the difficulties that arise from the fact that *every* market exchange depends upon some kind of "coercion." See especially 43 COLUM. L. REV. at 612-13; 38 POL. SCI. Q. at 478.

[50] See, for instance, the position of Bernard D. Broeker, who was a member of the subcommittees of the National Conference of Commissioners on Uniform State Laws and the American Law Institute, which considered the work on article 2 of the Code "leading to the 1958 Edition," UCC, comment to Title, at 8, and who remains at present a member of that subcommittee of the Permanent Editorial Board for the Uniform Commercial Code created to deal with possible changes in that article, *id.* at viii. At the meeting of the Larger Editorial Board of the American Law Institute on January 28, 1951, Mr. Broeker said (with respect to the 1950 version of § 2-302):

> I think that is a tremendous expansion of the law, and I don't think it ought to be that. I see no reason why I should not be allowed to make an unconscionable contract.

Proceedings of the Larger Editorial Board of the American Law Institute, January 27-28, 1951, at 172, partially reprinted in 6 BUS. LAW. 164, 184-85 (1951).

standing. The first clue is the express limitation of the unconscionability section to "form" contracts until 1948.[51] The second clue, which reinforces the first, is found by a study of the procedural unconscionability aspects of the ten cases which were inserted in the official comments to 2-302 in the 1950 draft,[52] described then (and now) as illustrating "the underlying basis" of the section,[53] and frequently pointed to by commentators in an effort to rebut any suggestion that the meaning of the unconscionability section is not clear.[54] As might have been expected, all ten of the cases involve commercial contracting situations. But they are notably common commercial situations. One finds none of the dramatics to be encountered in the equity unconscionability cases;[55] the parties are not notably old or young, bright or stupid, drunk, needy or sick. None of the cases involves sailors, women, heirs or other presumptive incompetents. There is no flavor of fraud or duress in any individualistic sense. Nothing more seems to have happened in any of these cases (except one) than that the parties entered into a contract on a pre-prepared form supplied by one of them.[56]

But one might have expected that more would have been disclosed about the contract-procuring procedure in these cases than merely that they involved form contracts. Unless all form contracts are to be deemed open to subsequent judicial rewriting under the unconscionability section, the cases illustrating the section's "underlying basis" should have given some indication of what, in addition to being printed, would make a contract vulnerable. When one examines the ten cases carefully, however, to see how many of them involved such factors, one finds, for instance, that less than half of them are merchant-consumer cases (though these would seem to furnish the best context for overreaching), the remainder being merchant-merchant trans-

[51] See 1948 DRAFT § 2-302.

[52] See 1950 DRAFT § 2-302, comment 1.

[53] See UCC § 2-302, comment 1. There has been absolutely no change in this portion of the comment from 1950 to date.

[54] See, *e.g.*, HAWKLAND, A TRANSACTIONAL GUIDE TO THE UNIFORM COMMERCIAL CODE § 1.1603, at 47-48 (1964); Davenport, *The Nebraska Uniform Commercial Code: An Introduction and Articles 1 and 2*, 43 NEB. L. REV. 671, 702 n.165 (1964); Lattin, *Article 2: Sales*, 23 OHIO ST. L.J. 185, 189 n.25 (1962). *But cf.* Note, 45 IOWA L. REV. 843, 849 (1960), suggesting that the official-comment cases are not quite on point.

[55] See text accompanying notes 185-202 *infra*.

[56] The clear exception is Austin Co. v. Tillman Co., 104 Ore. 541, 209 Pac. 131 (1922), where the governing contract was a letter composed for the occasion, by the party who ultimately successfully challenged its terms no less. The contracts in Kansas City Wholesale Grocery Co. v. Weber Packing Corp., 93 Utah 414, 73 P.2d 1272 (1937), Kansas Flour Mills Co. v. Dirks, 100 Kan. 376, 164 Pac. 273 (1917), and Robert A. Monroe & Co. v. Meyer, [1930] 2 K.B. 312, were uncomplicated enough to have been constructed for the particular transaction, but, although the cases are not explicit, it is more likely that they were printed forms.

actions.[57] Moreover, in none of the cases (except one) [58] does the court suggest that the form used in the transaction was particularly complicated, involved or extensive; these do not appear to be the kinds of forms that fight from ambush in a thicket of small print. Nor do the cases serve particularly to illustrate any monopolistic or oligopolistic power. Two of the cases [59] do indeed involve motor-vehicle warranty disclaimers of the kind eventually declared against public policy in *Henningsen v. Bloomfield Motors, Inc.*,[60] and another case [61] did involve a transaction between an automobile dealer and his supplier, a relationship replete with business-duress possibilities,[62] but that is as far as things go in this line.[63] These ten cases, then, are not a particularly good selection if they were chosen to illustrate varieties of commercial rapacity. These are, for the most part, simple form-contract cases, with no especially striking admixtures of quasi-fraud ("unfair surprise") or quasi-duress ("oppression"). It is inconceivable, therefore, that the cases are designed to establish a picture of what kinds of bargaining will cause the voiding of a contract or a clause therein. At best they may be taken to illustrate what kind of bargaining procedure will not serve to insulate a contract from gutting pursuant to 2-302 if it turns out to be substantively unconscionable. This is quite different from describing what bargaining conduct will

[57] Hardy v. General Motors Acceptance Corp., 38 Ga. App. 463, 144 S.E. 327 (1928), Bekkevold v. Potts, 173 Minn. 87, 216 N.W. 790 (1927), and Meyer v. Packard Cleveland Motor Co., 106 Ohio St. 328, 140 N.E. 118 (1922), are clear. In addition, the discontented seller in Kansas Flour Mills Co. v. Dirks, 100 Kan. 376, 164 Pac. 273 (1917), was apparently a farmer and one may include him among the nonmerchants even though it is arguable that a farmer selling his produce is a "merchant" under UCC § 2-104. See Corman, *The Law of Sales Under the Uniform Commercial Code*, 17 RUTGERS L. REV. 14, 17 (1962). *But see* 65 MICH. L. REV. 345 (1966).

[58] New Prague Flouring Mill Co. v. Spears, 194 Iowa 417, 438-39, 189 N.W. 815, 824 (1922): "if it be a contract, it is like the Apostle's conception of the human frame, 'fearfully and wonderfully made . . . .'" This is a merchant-to-merchant case, by the way.

[59] Hardy v. General Motors Acceptance Corp., 38 Ga. App. 463, 144 S.E. 327 (1928); Bekkevold v. Potts, 173 Minn. 87, 216 N.W. 790 (1927).

[60] 32 N.J. 358, 404, 161 A.2d 69, 95 (1960).

[61] Andrews Bros. v. Singer & Co., [1934] 1 K.B. 17 (C.A.).

[62] See Kessler, *Automobile Dealer Franchises: Vertical Integration by Contract*, 66 YALE L.J. 1135 (1957); Macaulay, *Changing a Continuing Relationship Between a Large Corporation and Those Who Deal With It: Automobile Manufacturers, Their Dealers, and the Legal System*, 1965 WIS. L. REV. 483, 740. It is worth noting that a statute now to some extent regulates the manufacturers' alleged gross power to overreach. 70 Stat. 1125 (1956), 15 U.S.C. §§ 1221-25 (1964).

I am assuming, just to state the case as strongly as possible, that the English automobile company in 1933 had as strong a stranglehold on its dealers as the American big four allegedly had on their dealers in the 1950's, but I would be exceedingly surprised if that were anything like the truth.

[63] For completeness one might include Kansas Flour Mills Co. v. Dirks, 100 Kan. 376, 164 Pac. 273 (1917), as a case where there might have been no real choice. I certainly do not know the state of Kansas wheat marketing in 1917, but I can conceive of an our-mill-or-none choice for a local farmer. There is no talk in the opinion of such a state of affairs, however.

void a contract; if these cases show the way, *any form contract* is up for grabs under 2-302.

The frank adoption of the position that any form contract was open to clause-by-clause policing, however, as Professor Llewellyn pointed out very early in the game, leaves this problem: the use of form contracts is a social good; it is the contracting-process component of the mass transaction, and the mass sales transaction has exceeding economic utility.[64] The form contract is designed *not* to be read or pondered; if it is or has to be it loses much of its utility. But not reading it leads to attempts at aggrandizement by form.[65] The law's problem, therefore, is to discourage dickering and overreaching simultaneously. For this it needs some new device, since in theory at least, until the time of the mass form, it was the dickering which discouraged the overreaching. If this new device, however, is making all printed forms open to after-the-fact *ad hoc* judicial second guessing, there is the danger that the efficiency of mass transactions will be seriously impaired. Moreover, once one faces the fact that the "vice" in the contracting process is nothing more than the use of a form contract, the internal justification for interfering with the parties' transaction becomes attenuated. It becomes exceedingly harder to justify suspension of the ordinary rule that a *sui juris* person who signs his name is bound to what is over his signature. After all, preprinting one's contracts is hardly *malum in se.*

This tension seems to have led some commentators on 2-302 to suggest that the contracting-procedure element which will permit scrutiny for unconscionability is not the mere use of a form but the use of a form *plus* something else. That is, they have felt impelled to find some "vice" to justify the judicial meddling.[66] And they have identified this form-plus situation with the "contract of adhesion," [67] a

[64] Llewellyn, Book Review, 52 HARV. L. REV. 700, 701 (1939).

[65] See MIMEO 1941 DRAFT § 1-C, comment A(2): "A private codification, however, has dangers. It may heap all the advantages sought on one side, and heap all the burdens on the other."

[66] See, *e.g.*, 1 HAWKLAND, A TRANSACTIONAL GUIDE TO THE UNIFORM COMMERCIAL CODE § 1.1602 (1964); Latty, *Sales and Title and the Proposed Code*, 16 LAW & CONTEMP. PROB. 3, 19 n.78 (1951).

[67] The term "contract of adhesion" most likely comes from Saleilles, who speaks of contracts:

> dans lesquels il y a la prédominance exclusive d'une seule volonté, agissant comme volunte unilatérale, qui dicte sa loi, non plus à un individu, mais à une collectivité indéterminée, et qui s'engage déjà par avance, unilatéralement, sauf adhésion de ceux qui voudrant accepter la loi du contrat, et s'emparer de cet engagement déjà crée sur soi-meme.

SALEILLES, DE LA DÉCLARATION DE VOLUNTÉ § 89, at 229-30 (1901).

Professor Patterson's translation of this passage is:

> in which a single will is exclusively predominant, acting as a unilateral will which dictates its law, no longer to an individual, but to an indeterminate collectivity, and which in advance undertakes unilaterally, subject to the

contract to which one of the parties must either "adhere" entirely or refuse altogether. In such a contract, a party may not bargain minutely over form or content, but must take it as is, if at all.[68] In some cases (for instance transactions with regulated utilities) a party may not even be able to bargain over price.[69] The essence of the adhesion contract is not its "formishness" [70] (that is just a symptom), but the fact that one of the parties has, at least for the purposes of the transaction in question, some of the powers of a monopolist. This "monopolistic power" need not be that wielded by a "true" monopolist, legal or other. It may be. One cannot bargain, especially over the price, with the telephone company, or with one's local airline. But that is not a requisite of the adhesion contract. In some cases the "monopoly" power may be only in a certain locality, when the purchaser is not mobile enough to get another seller who will offer other terms.[71] In some cases the monopoly power is really an expression of oligopoly power, *e.g.*, contract forms containing identical clauses written by competitors who nevertheless together blanket the market.[72] How it

adhesion of those who would wish to accept the law [loi] of the contract and to take advantage of the engagements imposed on themselves.

Patterson, *The Interpretation and Construction of Contracts*, 64 COLUM. L. REV. 833, 856 (1964). He goes on to theorize that the term contract of adhesion "may have been derived from the analogy of multilateral treaties, which are drawn up by negotiations between a few nations who sign and invite other nations to adhere to the treaty later." *Id.* at 856 n.96.

[68] The adhesion contract idea has had a long history of learned commentary, much of it of extremely high quality. See (in chronological order) Isaacs, *The Standardization of Contracts*, 27 YALE L.J. 34 (1917) (Karl Llewellyn was an editor of 27 YALE L.J.); PRAUSNITZ, THE STANDARDIZATION OF COMMERCIAL CONTRACTS IN ENGLISH AND CONTINENTAL LAW (1937), reviewed by Llewellyn, 52 HARV. L. REV. 700 (1939); Kessler, *Contracts of Adhesion—Some Thoughts About Freedom of Contract*, 43 COLUM. L. REV. 629 (1943); Ehrenzweig, *Adhesion Contracts in the Conflict of Laws*, 53 COLUM. L. REV. 1072 (1953); Sales, *Standard Form Contracts*, 16 MODERN L. REV. 318 (1953); Lenhoff, *Contracts of Adhesion and the Freedom of Contract: A Comparative Study in the Light of American and Foreign Law*, 36 TUL. L. REV. 481 (1962); Meyer, *Contracts of Adhesion and the Doctrine of Fundamental Breach*, 50 VA. L. REV. 1178 (1964); Wilson, *Freedom of Contract and Adhesion Contracts*, 14 INT'L & COMP. L.Q. 172 (1965).

[69] See, *e.g.*, 24 Stat. 384 (1887), as amended, 49 U.S.C. § 15 (1964) (railroad rates); Mo. ANN. STAT. § 392.240 (1952) (telephone rates).

[70] I am not sure whether the word "adhesion" has similar connotations in the original French, but it should not be overlooked that in English, perhaps because of the ubiquitous "adhesive tape," the word has strong connotations of agglutination and stickiness. This leads, I think, to an often subliminal impulse on the part of English-speaking commentators to speak of adhesion contracts as if they are not only quasi-monopolistic, but as if they are the kind that always contain a great many provisions stuck closely together. See, *e.g.*, 1 CORBIN, CONTRACTS § 128, at 552 (1963) ("long printed standardized contracts").

[71] It is of course a truism that there are geographic "relevant markets" within which a relatively small operator may function as a monopolist. But it should also be noted that the "market" sometimes is subjectively more narrow than that. For instance, consumers who are ignorant may think their local furniture stores are their only market, and this may give effective adhesion-contract power to the stores which they do not in fact have. *Cf.* Williams v. Walker-Thomas Furniture Co., 350 F.2d 445 (D.C. Cir. 1965), discussed at text accompanying notes 267-99 *infra*.

[72] *Cf.* Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960).

comes about is less important than the fact that it exists; the hallmark of the adhesion contract, and its alleged evil, is that the purveyor of such a contract is in the position for one reason or another to refuse to bargain, to put the other party to a take-it-or-leave-it option.

The dramatic situation which typically frames the contract of adhesion, therefore, is the merchant-consumer retail sale. But while it is very hard to imagine many adhesion contracts which are not at the same time form contracts, it is very simple to imagine form contracts which are not contracts of adhesion. In fact, there is one species of contract, one which most likely accounts for the bulk of commercial contracting in the nation, which is ordinarily a form contract but not an adhesion contract—the merchant-to-merchant form-pad contract, the subject matter of the "battle of the forms." [73] These form-pad deals may on occasion be adhesion deals too, but they certainly need not be. Indeed, there is often a sharp gulf between the typical contract of adhesion and the typical businessmen's battle of the forms. In a very large number of cases businessmen dealing with each other are not forced to take or leave each other's forms. They do not have so limited a market (or knowledge) that they cannot deal elsewhere, and they can, if they wish, argue about even the minutiae of the transaction.[74] As a general rule, however, they do not so wish. They prefer instead to maneuver like Renaissance condottieri for the cheap and bloodless positional victory that comes with the "making" of the contract on their own form. They acquiesce if they lose (if indeed they notice losing as such at all), seemingly because they just don't care. Admittedly a man does not make a contract expecting to get nothing. But he may very well enter a contract by which he assumes the *risk* of getting nothing. Professor Llewellyn with his customary insight noted that businessmen don't read contracts because they always expect to get the "something" they were dealing for. But he also stated as a fact that businessmen also expect to have subsidiary terms which are "fair," or at least "not manifestly unfair." [75] Let me submit an alternative possibility: most businessmen, insofar as they think about the question at all, expect that the other party's form will be the same

[73] See LLEWELLYN, THE COMMON LAW TRADITION 362-71 (1960). It may be seen in operation in Application of Doughboy Indus., Inc., 17 App. Div. 2d 216, 233 N.Y.S.2d 488 (1962).

[74] This is not to suggest that businessmen cannot be put into the position of adherers given proper leverage on the part of another businessman, as for instance may have occurred between automobile manufacturers and their dealers. See Kessler, *supra* note 62; Macaulay, *supra* note 62.

[75] See, *e.g.*, MIMEO 1941 DRAFT § 1-C(1)(d) and comment A(3); LLEWELLYN, *op. cit. supra* note 73, at 370 (1960) ("any not unreasonable or indecent terms"); Llewellyn, *Common Law Reform of Consideration: Are There Measures*, 41 COLUM. L. REV. 863, 871 (1941) ("too far unbalanced"); 21 AMERICAN LAW INSTITUTE PROCEEDINGS 114 (1944) ("a cake sliced 99-1").

kind of form which they had their lawyers draft for them, that is, a form which attempts to take everything for the owner of the pad. Does a man's own form have a warranty disclaimer? Yes. Does he expect to find one in the forms of his suppliers? You bet your life. Does he expect, therefore, that he will not receive the goods he ordered, of roughly the quality ordered? Absolutely not. He expects his supplier to deliver as per the order, just as he would expect to do to his customers, *but* he also expects that his supplier will have so drafted his form that it will be almost impossible to get legal recompense [76] if the buyer and seller disagree over the quality.

Professor Llewellyn may merely have known nicer businessmen than I. Perhaps his position is right as a matter of descriptive psychology. I suspect, however, that his "description" is really the prescription: businessmen *ought* not to try to take everything. That may well be true, but the evil is taking everything, not doing so by the presentation of a preprinted form contract. Insofar as anyone's justification for the utilization of a judicial-rewriting provision like 2-302 rests on the "evil" of the bargaining process, it is shattered if that evil consists only in prefabrication of the form itself; after all, the complaining party *could* have read it and if he had read it he *could* have argued about it. He didn't. It is at least arguable without a blush that such situations are to be treated in a fashion different from the treatment accorded true adhesion contracts [77] or even the standard quasi-monopolist consumer transaction.

But no distinction was made in 2-302 between merchant-to-merchant and merchant-to-consumer cases [78] (though the "merchant" definition was already made [79]). In addition, at least some of the commentary about various versions indicated that indeed the businessman's form pad was the target of the section, and it has certainly been a common assumption that none of the additional elements which

[76] Most of all, the businessman expects to settle things out of court and out of a law context. See Jones, *Merchants, the Law Merchant, and Recent Missouri Sales Cases: Some Reflections*, 1956 WASH. U.L.Q. 397, 411-18.

[77] For instance, (1) promulgating standard contract terms, see N.Y. INS. LAW § 155 (life insurance); Sales, *Standard Form Contracts*, 16 MODERN L. REV. 318, 340-42 (1953); Lenhoff, *Optional Terms (Jus Dispositivum) and Required Terms (Jus Cogens) in the Law of Contracts*, 45 MICH. L. REV. 39 (1946), and (2) setting up special tribunals to pass on form contracts in advance. See Gottschalk, *The Israeli Law of Standard Contracts, 1964*, 81 L.Q. REV. 31 (1965); Note, 66 COLUM. L. REV. 1340 (1966).

[78] It is interesting to remember that once upon a time there was a special rule for merchants' negotiations written into what was to become § 2-302. In 1943 DRAFT § 24 it was provided that if a merchant had an opportunity to read a contract, he was bound, even if he had not read it (the rule being otherwise for non-merchants). That provision was eliminated in MAY 1944 DRAFT § 23, never to reappear.

[79] See UCC § 2-104(1).

transmute a form contract into the somewhat more objectionable adhesion contract are prerequisite to the use of 2-302.[80] Thus, the use of the adhesion-contract learning is useful only insofar as it permits one to take the cachet and tone of the consumer-oppression cases and transfer it wholesale to 2-302. This enables one to feel that by using the section he is punishing naughty contracting conduct, without having to focus sharply on the fact that the level of conduct actually subject to 2-302 is hardly more than printing up one's contracts in advance.

Let us assume, however, that despite the references to the businessman's form-pad deal, the procedural unconscionability component of section 2-302 is at the adhesion-contract level rather than at the mere form-contract level; that is, that something more than mere preprinting must be shown before the resultant contract becomes subject to meddling under 2-302. It is exceedingly important to note that the only thing such a determination does is to set the level of contract-*insulating* conduct. One may now argue that a contract which has a sufficient number of indicia of compulsion to be fairly described as a contract of adhesion is not something upon which a party can rely to protect the provisions therein from the Code's unconscionability section. In other words, the adhesion contract becomes an exception to the usual rule that one is bound to that which he signs.[81] But that cannot mean that all contracts of adhesion are void, or that all clauses contained in contracts of adhesion are going to be stricken under 2-302. The presentation of an adhesion contract to a person is not, like the presentation of a pistol to his head, sufficient, if proven, to prevent the enforcement of the contract no matter how "fair" its terms.[82] The provisions of the telephone company tariffs and of the common carrier's tickets are ordinarily binding; one cannot get out from under a provision of that sort by showing only that one could not have bargained about it. Thus, once it is decided that a certain contract is vulnerable to scrutiny under 2-302 because its bargaining was not sufficiently angelic to insulate it from the section, the problem of the unconscionability provision of the Code still remains unsolved: granted that the contract is now open to 2-302, when is it, or a portion of it, "unconscionable"?

[80] See, *e.g.*, LLEWELLYN, *op. cit. supra* note 73, at 362-71; Llewellyn, *supra* note 75, at 869-70; Project, 10 U.C.L.A.L. REV. 1087, 1132 (1963); Note, 18 U. CHI. L. REV. 146, 146-47 (1950).

[81] RESTATEMENT, CONTRACTS § 70 (1932). RESTATEMENT (SECOND), CONTRACTS (Tentative Draft No. 1, 1964) seems not to change this "usual rule." Where old § 70 ought to be there is instead a reference to §§ 20-23. Of these, § 21(3) seems the most relevant, providing in effect that an only apparent assent may still be an assent, but that the resulting contract may be voidable because of "fraud, duress, mistake or other invalidating cause."

[82] See RESTATEMENT, CONTRACTS §§ 494-95 (1932). The same is true of fraud. See *id.* §§ 475-77.

SUBSTANTIVE UNCONSCIONABILITY

Drafting History

That the clause "or a portion of it" is necessary in the last sentence is particularly important for understanding the root failure of 2-302. For at least as originally conceived, substantive unconscionability meant something like "gross overall imbalance" of an entire contract. The section was viewed as governing unconscionable *contracts*, or at least large bloc segments of contracts drafted in gross. The determination invited by the section seems to have been almost quantitative. In the 1941 version of 2-302 and its extraordinarily extensive explanatory materials,[83] the metaphors of weighing and balancing abounded. The section described the act of which it was a part [84] as representing "a fair and balanced allocation of rights and liabilities between parties to sales and contracts to sell." [85] While "policy in general requires the parties' particular bargain to control," said the section,[86] and while "speed and convenience in transacting business may require . . . extensive departures [from the act's provisions] to be incorporated . . . ." [87]

> on the other hand, . . . where a group or bloc of provisions are not studied and bargained about in detail by both parties, then actual assent . . . is not in fact to be assumed where the group or bloc of provisions, taken as a whole, allocates rights and obligations in an unreasonably unfair and unbalanced fashion.[88]
>
> The policy of the legislature is also to avoid any unseeming portion of a bargain . . . under which one party seeks to displace the rules of this Act . . . in favor of a set of provisions which lack reasonable balance and fairness in their allocation of rights and obligations.[89]

When a number of matters are purportedly covered *en bloc* "as by a form contract," [90] the court may examine the bloc of provisions to see if it works a modification of the act's provisions "in an unfair and unbalanced fashion." [91] If, however, "the bloc as a whole is shown

[83] MIMEO 1941 DRAFT § 1-C. See note 13 *supra* for a description of the peculiarly discursive form of this original provision.

[84] At that time, the "Revised Uniform Sales Act." See MIMEO 1941 DRAFT title page.

[85] MIMEO 1941 DRAFT § 1-C(1)(a).

[86] MIMEO 1941 DRAFT § 1-C(1)(b).

[87] MIMEO 1941 DRAFT § 1-C(1)(c).

[88] MIMEO 1941 DRAFT § 1-C(1)(d).

[89] MIMEO 1941 DRAFT § 1-C(1)(e).

[90] MIMEO 1941 DRAFT § 1-C(2)(a).

[91] MIMEO 1941 DRAFT § 1-C(2)(a)(i).

affirmatively to work a fair and balanced allocation" it is not to be modified or stricken.[92] In addition, in "weighing fair balance" the court is directed "properly [to] consider the circumstances of preparation of any contract form . . . and, in particular . . . (ii) whether the displacement of the provisions of this Act sought by the form . . . as a whole runs disproportionately in favor of one party as against the other." [93]

This weighing-balancing, quasi-quantified outlook suggested by both the denotation and the metaphoric content of the statute is made even more explicit in the accompanying commentary:

> *The true principle is clear enough*: the expression of a body of fair and balanced usage is a great convenience, a gain in clarity and certainty, an overcoming of the difficulty faced by the law in regulating the multitude of different trades; on the other hand, the substitution of private rule-making by one party, in his own interest, for the balance provided by the law, is not to be recognized without strong reason shown.[94]

Or again:

> *Question for the court.* The total estimate of the effect of a body of provisions, in terms of balance, is a job for which a court is peculiarly fitted. The question of whether the provisions fit the circumstances of a particular trade is one which a special merchants' jury can best judge . . . . But the merchant runs some risk of accepting a provision merely as it is written because it is so written; and he has little training in sizing up a transaction from both ends at once, to reach a view of balance. As against this stands the fact that the *issue* to be tried is the issue of balance; and given that focus of attention, the merchant's jury would seem an adequate tribunal.[95]

Finally, the philosophical background of this decision to make the focus of the section the "imbalance" of the contract, was spelled out in the "Report" which accompanied the draft.[96] Under the heading, "The Problem of a Semi-Permanent Code of a Whole Field," the Report suggested that a Code could provide two kinds of statutory frameworks. The first kind (the example given was the Statute of Frauds) is "iron and unyielding; the parties must adapt themselves to it whether they will or no." But

---

[92] MIMEO 1941 DRAFT § 1-C(2)(a)(ii).

[93] MIMEO 1941 DRAFT § 1-C(2)(c).

[94] MIMEO 1941 DRAFT § 1-C, comment A(5), at 19.

[95] MIMEO 1941 DRAFT § 1-C, comment B(2). (Emphasis in original.) MIMEO 1941 DRAFT § 51-C made provision for the empanelling of a merchant's jury.

[96] See note 13 *supra*.

> the second kind of framework is a sort of standardized contract, serving wherever the parties have not particularized their bargain. It fills in and it fills out. Its office is to provide not only reasonable and fair solutions for particular matters, but, no less, a whole background of solutions for *any* matter, which as a whole is sufficiently reasonable and fair not to *need* to be bargained about.[97]

But, the Report goes on, variations from this "fair and balanced" background must be permitted if the admitted utility of form contracting is to be available at all under the statute. What then? Well, what the Report says about this balancing act so exposes the kernel of the problem that it bears full quotation:

> *Balance in any background sought to be substituted.*
>
> The Draft proceeds upon the assumption-in-policy that buyers and sellers ought (within the limits of such rules as those on legality) to be free to bargain as they choose. It proceeds upon the assumption-in-fact that choosing to bargain means resorting to deliberate and intentional dicker about particular terms, producing the kind of transaction known in law as an effective contract. Deliberate and intentional dickering is not shown in fact by a series of printed, unread clauses. When such a series appears, the position of the Draft is that the reasonableness of assuming *both* parties to have chosen and agreed to incorporate such a set of clauses, in silence and without dickering, depends upon whether the series of clauses presents the kind of balanced background which parties can fairly, or indeed accurately, be thought to incorporate by silence.[98]

This is developed in the comment to the withdrawn section 1-C. This passage is not only important for a study of this draft's position on procedural unconscionability, but it serves also to clarify the picture of the substantively unconscionable as viewed by the draftsman. "Imbalance," it seems clear, was not viewed only as evidence that some validating bargaining standard had not been met, but was also in itself that which was offensive ("unconscionable") in the resulting contract. Put another way, overall imbalance in the 1941 draft was not only evidence (perhaps proof) of procedural unconscionability; at the same time it was substantive unconscionability.

The 1941 draft, then, its comments and accompanying "Report," with their complete focus upon overall imbalance, must have contem-

[97] NATIONAL CONFERENCE OF COMMISSIONERS ON UNIFORM STATE LAWS, PROPOSED REPORT ON AND DRAFT OF A REVISED UNIFORM SALES ACT 22-23 (1941). (Emphasis in original.)

[98] *Id.* at 24. (Emphasis in original.)

plated as the field of operation of 2-302 the entire contract, or at least a major group of provisions within the contract.[99] "Imbalance" is, to put things mildly, a singularly inartistic way to refer to what might be objectionable about a single contractual provision.[100] And indeed, after this first version of the unconscionability section was withdrawn from consideration as "unworkable," [101] the provision which replaced it made even more explicit that the unconscionability decision was to be made with respect to the whole contract. The new section read as follows:

> SECTION 24. FORM CLAUSES, CONSCIONABLE AND UNCONSCIONABLE.
>
> (1) A party who signs or accepts a writing evidencing a contract for sale which contains or incorporates one or more form clauses presented by the other party is bound by them unless the writing *when read in its entirety* including the form clauses is an unconscionable contract . . . .[102]

The accompanying comment [103] made it even more abundantly clear that the vice still being attacked was lack of overall contractual balance. Referring to form-pad transactions, the comment stated that "such forms when drawn with *elaborate lopsidedness* can become what are in essence instruments of trickery." [104] And with regard to what the final official comment should say, the unofficial comment suggested:

> The Comment should show that since the rules of the Act are drawn with a careful balance of the rights and needs of buyer and seller, a form which cumulates too many departures from those rules in material particulars, and in favor of one side only, begins to take on the aspect of the unconscionable.[105]

---

[99] While I think that the truth of this conclusion is established beyond cavil, it should be pointed out that the heading to § 1-C read, "Declaration of Policy and Procedure with Regard to Displacement of Single Provisions or Groups of Provisions by Agreement."

[100] It is of course possible that in some circumstances a single provision of an entire contract might be so outrageous as to render the whole radically unbalanced. For instance, if *no* duties are given one of the parties, even at common law this absolute imbalance prevented enforcement under the rubric "illusory contract." See 1 CORBIN, CONTRACTS § 145 (1963). A too enthusiastic limitation-of-remedy clause might have a similar effect under the Code. *Cf.* UCC § 2-719(2).

[101] See note 21 *supra*.

[102] 1943 DRAFT § 24. (Emphasis added.)

[103] [Llewellyn,] Informal Appendix To Revised Uniform Sales Act, Third Draft, 1943, Tentative Sketch of Material for Comments (1943).

[104] *Id.* at 11. (Emphasis added.)

[105] *Id.* at 12.

This idea, that "unconscionability" meant something like overall contractual imbalance, was maintained all the way up to the 1948 version of section 2-302.[106] At that point came a change of immense significance. The 1948 version read in its entirety as follows:

> SECTION 23. UNCONSCIONABLE CONTRACT OR CLAUSE.
>
> (1) If the court finds the contract to be unconscionable, it may refuse to enforce the contract or strike any unconscionable clauses and enforce the rest of the contract or substitute for the stricken clause such provision as would be implied under this Act if the stricken clause had never existed.
>
> (2) A contract not unconscionable in its entirety but containing an unconscionable clause, whether a form clause or not, may be enforced with any such clause stricken.

This draft says bluntly that a court may excise from a not unconscionable contract any single "unconscionable" clause, and the comment accompanying the 1949 version (in which version the text of the section itself is not changed from the 1948 draft) says it just as bluntly:

> Under this section the court, in its discretion, may refuse to enforce the contract as a whole if it is permeated by the unconscionability or it may strike any single clause or group of clauses which are so tainted or which are contrary to the essential purpose of the agreement.[107]

From this point on in the drafting history of section 2-302 the concept of single-clause unconscionability was fixed; no substantial changes were made in this regard in the text of the statute or its accompanying comments.[108] The current comment 2[109] was present in essentially its final form as early as the 1950 draft.[110]

This progression through the drafts of the idea of substantive unconscionability, from overall imbalance to one-clause naughtiness, is the most important single transformation disclosed by a study of the drafting history. Determining which contracts are substantively unconscionable is a difficult enough job even if one's conception of substantive unconscionability is something like "gross imbalance" or

---

[106] 1948 DRAFT § 23.

[107] 1949 DRAFT § 2-302, comment 1. There were no comments accompanying 1948 DRAFT.

[108] See note 39 *supra*.

[109] 2. Under this section the court, in its discretion, may refuse to enforce the contract as a whole if it is permeated by the unconscionability, or it may strike any single clause or group of clauses which are so tainted or which are contrary to the essential purpose of the agreement, or it may simply limit unconscionable clauses so as to avoid unconscionable results.
UCC § 2-302, comment 2.

[110] Compare 1950 DRAFT § 2-302, comment 2.

"lopsidedness." After all, even a lopsided contract might in some cases be hard to identify; what if X got seven risks and Y got five—or four to three—or two to one? The spuriousness of the quantification lurking in the idea of a contract suffering from "overall imbalance" is a potential plague for close cases. Compared, however, with the difficulties of dealing with a concept of one-clause unconscionability, it is pure vanilla. It is not a rare rule in the law that he who bites off much more than he should will be judicially choked. The treatment of over-enthusiastic no-competition clauses [111] and trade-secret protection provisions [112] are common examples. Quantification, while falsifying if it gives the impression of numerical precision, at least carries with it a metaphorical framework which is an aid to the decision of all but the closest cases. A real scale is admittedly useless for the measurement of anything but physical weight, but a scale as a metaphor at least lets one know that he is looking for too much of something. Admittedly, the precision of the result depends upon what is being "weighed." If it is something like potatoes which arbitrarily have a "weight," then the measure of weight *in those terms* is exact, but if it is a quality not attracted by gravity which is "weighed," then the weighing and balancing are not going to be more than metaphorically precise. Risks, for instance, do not have calibratable weight. Once it is established, however, that one is looking for a comparison of risks, if what is involved is a contract which gives no risks at all to one of the parties, or almost none, the decision under the metaphor is easy. And it was this sort of contract which seems to have been the intended target of the original draftsman's original draft.[113] For that "almost-all" kind of contract

[111] See, *e.g.*, Purchasing Associates, Inc. v. Weitz, 13 N.Y.2d 267, 196 N.E.2d 245 (1963), for a recent decision, the four-to-three nature of which is eloquent on how hard these determinations can be. See also Vaughan v. Kizer, 400 S.W.2d 586, 589 (Tex. Civ. App. 1966) (reasonableness of restrictions as "questions of law for determination by the court"); Brown v. Devine, — Ark. —, 402 S.W.2d 669 (1966) (must strike whole contract; "modifying" not within court's power).

[112] See Note, *Protection of Inventive Ideas Through Postemployment Assignment Covenants*, 1965 WASH. U.L.Q. 335, 359.

[113] See text accompanying notes 83-98 *supra*.

Of course, nothing as difficult as this question is ever that easy for a draftsman like Llewellyn. In 1944 he told the American Law Institute:

> I think that everybody who signs up on such a form knows perfectly well that he is signing a contract drawn to some extent in favor of the other party and against him, and he is perfectly willing to take a cake sliced 60-40 or perhaps even 75-25. But when it gets to be a cake sliced 99-1, he doesn't find that that is what he was agreeing to tacitly.

21 AMERICAN LAW INSTITUTE PROCEEDINGS 114 (1944). Ten years later (and, it should be noted, *after* the central concept of the section had switched to one-clause unconscionability), his metaphor was still going strong (with only slight gustatory variation); it was then "80% of the pie." NEW YORK LAW REVISION COMM'N, STUDY OF THE UNIFORM COMMERCIAL CODE 113 (Legis. Doc. No. 65(B), 1954). On the other hand he was also sometimes plagued by a feeling that certain *individual* clauses were just no good, no matter how well bargained, *e.g.*, no-oral-waiver clauses (other than between merchants). Llewellyn, *Common-Law Reform of Consideration: Are*

ordinarily involved in a form contract, and a fortiori in an unregulated contract of adhesion, this rough quantitative approach seems to have been roughly sufficient.

The overall-imbalance formulation, however, cannot settle all of the heart stirrings which may be caused by harsh results. If all of the risks of a particular contract are put upon A, except for one which is put upon B, and it is the risk which B was to bear which in fact occurs, B loses. Under the overall-imbalance rubric how does one deal, for instance, with an ordinary old contract with no radical clauses of any kind, which, nevertheless, contains a clause clearly disclaiming any warranty? The problem may be stated quite simply: with respect to the effect of any particular contract upon any actual party thereto, *most of the contract is irrelevant.* Ordinarily only one shifted risk comes home to roost, and if there were fifty others shifted, their potential is never actualized. That means that a refusal to enforce a contract in any particular case would not be a response to what happened in that case (which would have happened anyway had only the one risk which came true been shifted), but would be instead a response to general naughtiness on the part of the party who procured such a tough contract.[114]

On the other hand, if one decides to police contracts on a clause-by-clause basis, he finds that he has merely substituted the highly abstract word "unconscionable" for the possibility of more concrete and particularized thinking about particular problems of social policy. Should warranty disclaimers be permitted? If so, should they be with respect to consumer goods?[115] Should parties be allowed to agree about what law will govern their contract?[116] To what extent, if any, should a party be permitted to limit his liability under a contract? All of these questions need decision.[117] *But not one of them is helped*

There Measures?, 41 COLUM. L. REV. 863, 869 (1941). And even his latest writings show that the existence of an important one-clause-whole-contract distinction was not totally appreciated. See LLEWELLYN, THE COMMON LAW TRADITION 371 (1960) (contract terms should be unfair "neither in the particular nor in the net").

[114] An instance of this peculiar type of decision is Campbell Soup Co. v. Wentz, 172 F.2d 80 (3d Cir. 1948), where the defendant willfully broke an unfavorable contract but the plaintiff was denied specific performance because of the alleged nastiness of certain provisions in the form contract he drafted, even though they had nothing to do with the defendant's breach or impending loss.

[115] See Franklin, *When Worlds Collide: Liability Theories and Disclaimers in Defective-Product Cases*, 18 STAN. L. REV. 974, 1019 (1966), for just such more concretized thinking.

[116] See the exchange between Beutel, *The Proposed Uniform [?] Commercial Code Should Not Be Adopted*, 61 YALE L.J. 334, 350-52 (1952), and Gilmore, *The Uniform Commercial Code: A Reply to Professor Beutel*, 61 YALE L.J. 364, 373-74 (1952).

[117] And they all do get some kind of particularized decision in the Code. UCC § 2-316 is on exclusion of warranty, § 1-105 is on choice of law and § 2-719 is on limitation of remedy. It is in fact hard to imagine which kinds of clauses reasonably expectable in a commercial contract might be unconscionable but have not been regulated by more specific portions of the Code.

toward solution by being subsumed in a section as a species of "unconscionability." The word "unconscionable," as finally used in the Code, describes neither the dramatic situation of two persons bargaining nor the "imbalance" or "lopsidedness" or other quality of the resulting contract, but rather describes the emotional state of the trier which will justify his use of the section. In other words, the attitudes relevant under section 2-302 are not those of the parties but those of the judges. The pictures to be sought in the facts are not of the varieties of oppressive or surprising negotiations, nor of oppressive or surprising contracts, but rather of oppressed or surprised judges. But what may permissibly make the judges' pulses race or their cheeks redden, so as to justify the destruction of a particular provision, is, one would suppose, what the judge ought to have been told by the statute. In short, once the movement was made to a conception of one-clause unconscionability, and the "overall-imbalance" rubric was abandoned as insufficient, the statute and its commentary had been stripped of any power to guide the decision of what the "bad" single provisions might be like. And the enormous significance of this failure may be illustrated by a careful consideration of the ten cases described as disclosing the "underlying basis" of the section,[118] and the interesting way they failed to fill the gap.

The Official-Comment Cases

The ten cases do illustrate the one-clause-unconscionability theory, each really involving only one offensive "unconscionable" clause. But there are only two types of naughty clauses represented: warranty disclaimers and remedy limitations.[119] Given this arresting fact alone, one might be tempted to conclude that the purpose of section 2-302 was to render warranty disclaimers and remedy limitations per se unconscionable. Nor indeed would the world, even the commercial world, come to an end if parties were forbidden either to disclaim warranties or to withhold from each other any of the total panoply of remedies for breach of contract which the Code provides.[120] In other places in the Code the draftsmen have felt free flatly to forbid particular contractual provisions, agreement between the parties or not.[121]

[118] UCC § 2-302, comment 1.

[119] The cases appear to be divided about half and half. See HONNOLD, CASES ON SALES AND SALES FINANCING 27 (2d ed. 1962). In several cases it is hard to tell if the clause at issue is better classified as a disclaimer of warranty or as a limitation of remedy should a warranty be found to be breached. *E.g.*, Robert A. Munro & Co. v. Meyer, [1930] 2 K.B. 312, 314 ("the goods to be taken with all faults and defects; damaged or inferior, if any, at valuation to be arranged mutually or by arbitration").

[120] See UCC §§ 2-702-17.

[121] For instance, reducing the limitation period to less than one year is forbidden by UCC § 2-725(1).

Unfortunately for the solution of the problem now before us, that road was not the one taken. It is perfectly clear that under the Code warranties may be disclaimed, and remedies for breach may be modified and limited; neither are per se unconscionable. Section 2-316 of the Code is devoted to describing the procedure to be used in disclaiming warranties, and section 2-719 is devoted to doing the same job for remedy limitations. If, therefore, the substantive provisions of the contracts involved in the ten official-comment cases are to have any bearing upon the definition of substantive unconscionability, one must discover if and to what extent those two types of clauses might comply with their own particularized sections and yet fall afoul of section 2-302, or find some analogical model which will make the two provisions descriptive of the *kind* of provisions being aimed at.

The simpler case of the two is presented by the remedy-limitation problem. Section 2-719 of the Code provides as follows:

> (1) Subject to the provisions of subsections (2) and (3) of this section and of the preceding section on liquidation and limitation of damages,
>
> (a) the agreement may provide for remedies in addition to or in substitution for those provided in this Article and may limit or alter the measure of damages recoverable under this Article, as by limiting the buyer's remedies to return of the goods and repayment of the price or to repair and replacement of non-conforming goods or parts; and
>
> (b) resort to a remedy as provided is optional unless the remedy is expressly agreed to be exclusive, in which case it is the sole remedy.
>
> (2) Where circumstances cause an exclusive or limited remedy to fail of its essential purpose, remedy may be had as provided in this Act.[122]

Its first official comment reads:

> Purposes:
>
> 1. Under this section parties are left free to shape their remedies to their particular requirements and reasonable agreements limiting or modifying remedies are to be given effect.
>
> However, it is of the very essence of a sales contract that at least minimum adequate remedies be available. If the parties intend to conclude a contract for sale within this

[122] UCC §§ 2-719(1), (2).

> Article they must accept the legal consequence that there be at least a fair quantum of remedy for breach of the obligations or duties outlined in the contract. Thus any clause purporting to modify or limit the remedial provisions of this Article in an unconscionable manner is subject to deletion and in that event the remedies made available by this Article are applicable as if the stricken clause had never existed. Similarly, under subsection (2), where an apparently fair and reasonable clause because of circumstances fails in its purpose or operates to deprive either party of the substantial value of the bargain, it must give way to the general remedy provisions of this Article.[123]

Now, what is most striking is the extent to which the "evil" with which this section and its comment are to deal is *not* left in terms of high-level abstraction. First, the question of procedural conscionability is not ambiguously left hanging. The section makes clear that even if it were proved by the proverbial twenty eavesdropping bishops that a particular remedy limitation had been haggled over between the parties, the limitation would have to go if "circumstances" deprived a party of "the substantial value" of his bargain. No matter what actual bargaining had led to the remedy-limitation, "at least minimum adequate remedies" must be provided in the contract.

Thus, the Code reflects a substantive decision on this point. It did not say that remedies for breach could not be limited to less than those provided in the Code, but it did provide that remedy for breach could not be *eliminated* by agreement. Certainly, section 2-719 did not settle all of the problems, most particularly what a "minimum adequate remedy" might be.[124] But it did settle the question to the extent of providing that no remedy at all was in fact below that requisite minimum. In other words, the "unchangeable background" view which animated much of earlier drafts of the Code[125] and Karl Llewellyn's thinking from a period even before the drafting began,[126] is to some small extent preserved in section 2-719's attack. Put still another way, section 2-719 represents a drafting decision that at least one form of gross overall imbalance will not be permitted. This approach accords

---

123 UCC § 2-719, comment 1.

124 See Note, *Limitations on Freedom To Modify Contract Remedies*, 72 YALE L.J. 723 (1963), for a recent discussion of the general problem.

125 1950 DRAFT § 1-107 provided: "The rules enunciated in this Act which are not qualified by the words 'unless otherwise agreed' or similar language are mandatory and may not be waived or modified by agreement." By the 1952 DRAFT that provision had disappeared and in the current version of the Code the power to modify by agreement has been made explicit. See UCC §§ 1-102(3), (4).

126 See, *e.g.*, Llewellyn, *On Warranty of Quality, and Society: II*, 37 COLUM. L. REV. 341, 403-04 (1937).

with the Code's section dealing with liquidated damages clauses,[127] which provides in an orthodox way[128] that too much in the way of damages for breach is a voidable "penalty." Its comment then added that, "An unreasonably small amount . . . might be stricken under the section on unconscionable contracts or clauses."[129] Section 2-719 merely moves the argument one reasonable step further: if one cannot provide too little in the way of liquidated damages, one cannot provide too little in the way of modes of recourse either.

The standard provided by section 2-719, therefore, is also a quasi (or spuriously) quantitative one: is there *some* remedy provided; if so, is the remedy "enough"? This is hardly a universal solvent for all of the problems that might arise in this area. But it is hard to think of any factual situation in which asking oneself whether a provision were "unconscionable" would clarify the decision to a problem left unsolved after asking oneself the much more particularized questions suggested in section 2-719. As benchmarks for determining the permissibility of a remedy limitation, 2-302's "oppression and unfair surprise" can't hold a candle to 2-719's "fail of its essential purpose," "minimum adequate remedy," and "fair quantum of remedy." Obviously the 2-719 catchwords don't make close cases easy, but they certainly do a better job than the single word "unconscionability." It is as if a single statute contained two provisions, one which forbids the charging of "excessive" interest and another forbidding "lender naughtiness." Neither of these sections would be much help in settling what to do with a 7½% interest charge. But if the interest rate were, say, 78%, one could handle the problem pretty easily with the excessiveness section; it is hard to see what the naughtiness section would add. In brief, when two sections deal with the same conduct, and one deals particularistically with reasonably clear standards, and the other deals with the problem only in terms of emotional coloration, the latter provision is unlikely to be of any help in solving a problem of specific application.[130]

---

[127] UCC § 2-718(1).

[128] See 5 CORBIN, CONTRACTS §§ 1054-75 (1964); RESTATEMENT, CONTRACTS § 339 (1932).

[129] UCC § 2-718, comment 1.

[130] This is put into relief, I think, by what happened to § 2-719 itself when it ceased to be particularistic. Section 2-719(3) reads as follows:

> (3) Consequential damages may be limited or excluded unless the limitation or exclusion is unconscionable. Limitation of consequential damages for injury to the person in the case of consumer goods is prima facie unconscionable but limitation of damages where the loss is commercial is not.

The official comment directed to this subsection contains three sentences, one tautological, one truistic and one mysterious, as follows:

> 3. Subsection (3) recognizes the validity of clauses limiting or excluding consequential damages but makes it clear that they may not operate in an

The mysterious last sentence in the third comment to section 2-719 [131] says that the "seller *in all cases* is free to disclaim warranties in the manner provided in Section 2-316." [132] If that in fact means what it seems to, that so long as the procedure set forth in section 2-316 is followed any warranty may be disclaimed, then the significance of the official-comment cases to an understanding of what might be substantively unconscionable is even more shadowy. Since the official-comment cases which do not deal with liability limitations deal instead with warranty disclaimers, if section 2-302 is inapplicable to warranty disclaimers too it is hard to see that the contract clauses involved in the official-comment cases could have much bearing on a definition of substantive unconscionability.

As I suggested earlier, it would not have been inconceivable for the draftsmen simply to have declared that some or all of the traditional implied warranties surrounding sales would be nondisclaimable.[133] They did not do so. The section of the Code explicitly devoted to the problem of warranty disclaimer is much more a blueprint of disclaiming technique than an extended form of interdiction. Section 2-316 of the Code provides (in its relevant portions) as follows:

> (2) Subject to subsection (3), to exclude or modify the implied warranty of merchantability or any part of it the language must mention merchantability and in case of a writing must be conspicuous, and to exclude or modify any implied warranty of fitness the exclusion must be by a writing and conspicuous. Language to exclude all implied warranties of fitness is sufficient if it states, for example, that "There are no warranties which extend beyond the description on the face hereof."
>
> (3) Notwithstanding subsection (2)
>
> (a) unless the circumstances indicate otherwise, all implied warranties are excluded by expressions

---

unconscionable manner. Actually such terms are merely an allocation of unknown or undeterminable risks. The seller in all cases is free to disclaim warranties in the manner provided in Section 2-316.

The central difficulty in this provision is that *all* consequential-damage exclusions operate in the *same* manner, and that is a very harsh manner indeed, *viz.*, he who has suffered a consequential loss does not get compensated for it. Since § 2-719 applies to substantively offensive clauses whether bargained about or not, the question cannot turn on procedural unconscionability considerations. What then does it mean that something is "prima facie" unconscionable? Is that a statement about the burden of going forward at a trial, or the burden of persuasion, or both? Or is it just a quiet way of saying that it is in all cases unconscionable? The awkwardness, I suspect, is the result of trying to give content to the shibboleth "unconscionable," instead of saying flat out what was meant.

[131] See note 130 *supra*.

[132] UCC § 2-719, comment 3. (Emphasis added.)

[133] Perhaps some provision would have had to have been made for those rare as-is sales, for instance, jalopies to teenagers, but that could easily have been handled by a more explicit version of present § 2-316(3)(a).

> like "as is," "with all faults" or other language which in common understanding calls the buyer's attention to the exclusion of warranties and makes plain that there is no implied warranty; and
>
> (b) [as to patent defects] when the buyer before entering into the contract has examined the goods . . . as fully as he desired or has refused to examine the goods . . . and
>
> (c) an implied warranty can also be excluded or modified by course of dealing or course of performance or usage of trade.[134]

In case anyone could still doubt the disclaimability of warranties, the comments accompanying 2-316 provide, *inter alia* as follows:

> 1. This section is designed principally to deal with those frequent clauses in sales contracts which seek to exclude "all warranties express or implied." It seeks to protect a buyer from unexpected and unbargained language of disclaimer by denying effect to such language when inconsistent with language of express warranty and permitting the exclusion of implied warranties only by conspicuous language or other circumstances which protect the buyer from surprise.
>
> . . . .
>
> 3. Disclaimer of the implied warranty of merchantability is permitted under subsection (2), but with the safeguard that such disclaimers must mention merchantability and in case of a writing be conspicuous.
>
> 4. Unlike the implied warranty of merchantability, implied warranties of fitness for a particular purpose may be excluded by general language, but only if it is in writing and conspicuous.[135]

Section 2-316, then, not only says that warranties may be disclaimed, but it says how one should go about doing so, in rather impressive detail and with surprising particularity. It is obvious that

[134] UCC § 2-316. Subsection (1) of § 2-316 deals with the conflicts between express warranties and disclaimers, saying that the warranty and the disclaimer "shall be construed whenever reasonable as consistent with each other" but that "negation or limitation is inoperative to the extent that such construction is unreasonable." What, if anything, that subsection might mean is, for obvious reasons, the subject of some dispute. See Ezer, *The Impact of the Uniform Commercial Code on the California Law of Sales Warranties*, 8 U.C.L.A.L. REV. 281, 310-311 (1961); Note, 112 U. PA. L. REV. 564, 581 n.145 (1964). Subsection (4) of § 2-316 is a cross-reference provision pointing to §§ 2-718 and 2-719, liquidation and limitation of damages sections, respectively. It does not do any pointing to § 2-302.

[135] UCC § 2-316, comments 1, 3, 4. See also UCC § 2-315, comment 6 (which deals with the implied warranty of fitness for a particular purpose): "The specific reference forward in the present section . . . is to call attention to the possibility of eliminating the warranty in any given case."

the vice is "surprise," and thus even the word "conspicuous" at the very heart of the provision is not left to speculation. Section 1-201(10) of the Code, which defines "conspicuous" generally as "so written that a reasonable person against whom it is to operate ought to have noticed it," goes on thereafter to simplify matters by incorporating a short typographic manual for conspicuousness: "A printed heading in capitals (as: NON-NEGOTIABLE BILL OF LADING) is conspicuous. Language in the body of a form is 'conspicuous' if it is in larger or other contrasting type or color." [136]

With these standards before us, let us now test a case under the Code's disclaimer-of-warranty provision. To make it a fair test, let us make one of the personae a consumer,[137] and, in fact, choose a dramatic situation suggested by the Code itself.[138] Sir Edmund Pillory, an eminent mountain climber, enters Abercrombie & Fitch in New York to buy some shoes. The salesman, recognizing Sir Edmund immediately, rushes over to serve him. Sir Edmund orders "some good sturdy shoes" and the salesman, knowing that they are wanted for Sir Edmund's highly advertised impending climb up K-3, brings out a pair which, while fine for walking upon ordinary ground, is hardly sufficient for mountain climbing. Sir Edmund purchases the shoes, and thereafter makes it only to roughly K-2½. His death is at least arguably attributable to the inappropriate shoes. His executor sues Abercrombie & Fitch for breach of warranty of fitness for a particular purpose. The store defends on the ground that on the sales form handed to Sir Edmund at the time he purchased the shoes there was the following form statement in red (a contrasting color): "There are no warranties which extend beyond the description on the face hereof." That is, the disclaimer was put in the exact language specified in section 2-316(2) as being an incantation sufficient for this particular purpose, and it is in a form defined by the Code as "conspicuous." In fact, Sir Edmund never read the disclaimer, and had he done so, being a man of action rather than words, he would not have understood it. It would appear, nevertheless, that on these facts, the requirements of the Code for the successful disclaimer of a warranty having been expressly met, Messrs. Abercrombie & Fitch would be home free under 2-316, and that this would be true even though Sir Edmund's claim is

[136] UCC § 1-201(10). The same provision makes conspicuousness seem even less a matter of degree in the draftsmen's eyes by making it a question for the court rather than the jury.

[137] *Cf.* UCC § 2-719(3) (prima facie unconscionability of consumer-goods remedy limitation).

[138] See UCC § 2-315, comment 2, an attempt to flesh out the exact meaning of "particular purpose": "For example, shoes are generally used for the purpose of walking upon ordinary ground, but a seller may know that a particular pair was selected to be used for climbing mountains."

for "injury to the person in the case of consumer goods," the limitation of damages with respect to which would have been "prima facie unconscionable" under 2-719(3). Does the matter end there, however, or may 2-302 nevertheless be applied to this state of facts to eliminate the effect of the disclaimer as "unconscionable"? In other words, given a careful meeting of the requirements of section 2-316 (the Code section particularly devoted to warranty disclaimers), may the more generally protective and loosely defined section devoted to general naughtiness be invoked to avoid the harsh result?

Almost everyone seems to think so. It appears to be a matter of common assumption that section 2-302 *is* applicable to warranty disclaimers.[139] I find this, frankly, incredible. Here is 2-316 which sets forth clear, specific and anything but easy-to-meet standards for disclaiming warranties. It is a highly detailed section, the comments to which disclose full awareness of the problem at hand. It contains no reference of any kind to section 2-302, although nine other sections of article 2 contain such references.[140] In such circumstances the usually bland assumptions that a disclaimer which meets the requirements of 2-316 might still be strikable as "unconscionable" under 2-302 seems explainable, if at all, as oversight, wishful thinking or (in a rare case) attempted sneakiness.[141]

Of course, the emotional pressure to reach a no-disclaimer result via the unconscionability route if it cannot be done otherwise is understandable. One need only point out that if in the *Henningsen* case [142] the auto manufacturers had gotten together to agree upon a form of disclaimer clause which accorded with the requirements of section 2-316, under my view, Mrs. Henningsen's serious personal injuries would have to go uncompensated. This would be so even though the auto manu-

[139] Just to list some of the clearest cases, CONTINENTAL ILLINOIS NATIONAL BANK & TRUST COMPANY OF CHICAGO, SUMMARY OF THE UNIFORM COMMERCIAL CODE FOR ILLINOIS 8 (1961) ("will perhaps find its most frequent use"); Cudahy, *Limitation of Warranty Under the Uniform Commercial Code*, 47 MARQ. L. REV. 127, 128-29 (1963); Duesenberg, *The Manufacturer's Last Stand: The Disclaimer*, 20 BUS. LAW. 159, 162 (1964); Lauer, *Sales Warranties Under the Uniform Commercial Code*, 30 MO. L. REV. 259, 283 (1965) ("would apply, of course"); Weeks, *The Illinois Uniform Commercial Code: Article 2—Sales*, 50 ILL. B.J. 494, 516 (1962); Note, 45 IOWA L. REV. 843, 857 (1960) ("All contractual provisions are subject"); Note, 109 U. PA. L. REV. 401, 420 (1961). See also Peters, *Remedies for Breach of Contracts Relating to the Sale of Goods Under the Uniform Commercial Code: A Roadmap for Article Two*, 73 YALE L.J. 199, 282 (1963), recognizing the need for "a sufficiently motivated court." A case taking the position clearly (but in dictum) is Willman v. American Motor Sales Co., 44 Erie Leg. J. 51, 57 n.3 (Erie County Ct. Pa. 1961). *But cf.* Franklin, *supra* note 115, at 994-95, 1013-14, for a more skeptical position.

[140] UCC §§ 2-202, 2-204, 2-205, 2-207, 2-303, 2-508, 2-615, 2-718, 2-719.

[141] In this last category see Note, 43 B.U.L. REV. 396, 403-04 (1963), a disingenuous (or ingenious) attempt to suggest that if the comments to the section indicate that the section would not apply, the comments, since not "part of" the statute, ought to be disregarded.

[142] Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960).

facturers had the oligopolistic power to make the terms of their contracts unbargainable. Any court might find it intolerable to allow a rich auto manufacturer to avoid making restitution for injuries suffered through the breakdown of a dangerous instrument it manufactured merely because it had made verbal compliance with a talismanic form of words which may not have been read or understood by the purchaser,[143] and about which he could have done nothing even if he had read, understood and objected. Such a decision would be an exceedingly painful one to announce. But that *is* what the statute says. There is nothing to prevent a legislature from regulating certain particular contractual provisions out of existence, as they have done on innumerable occasions in the past.[144] Certainly there is not much force remaining in simplistic freedom-of-contract arguments that legislatures may not determine, as a matter of policy, that some things in contracts just won't go.[145] The Code itself goes that route in other places [146] and there would have been nothing offensive in doing so with respect to warranty disclaimers, especially with respect to consumer goods. What is offensive is the seeming attempt on the part of some commentators to nullify the legislative determination that warranty disclaimers, for the time being at least, may continue.[147] Even legislatures, one would think, are entitled to some protection from oppression and unfair surprise.

---

[143] The court in *Henningsen* might and may have relied upon the "hidden-provision" argument alone to reach its result. See 74 HARV. L. REV. 630, 631 (1961).

[144] For instance, for a whole list of such statutory prescriptions applicable specifically to the consumer-contract field, see CURRAN, TRENDS IN CONSUMER CREDIT LEGISLATION 311-22 (1965).

[145] That government has such power is today accepted almost as a postulate, see, *e.g.*, GOODHART, ENGLISH LAW AND THE MORAL LAW 109 (1953); RADCLIFFE, THE LAW AND ITS COMPASS 65-66 1960), the powerful and successful attacks on overenthusiastic interpretations of "freedom of contract" having come a long time ago. See, *e.g.*, Lochner v. New York, 198 U.S. 45, 74-76 (1905) (Holmes, J., dissenting); Hamilton, *Freedom of Contract*, 3 ENCY. SOC. SCI. 450 (1931). That does not mean that the attractions of that "freedom" are wholly gone. *Compare* Pound, *Liberty of Contract*, 18 YALE L.J. 454 (1909), *with* Pound, *Law in the Service State: Freedom Versus Equality*, 36 A.B.A.J. 977, 1050-53 (1950).

The feeling that there once was a time when men's promises were more trustworthy is not a new one. See Chaucer, *Lak of Stedfastness* (ca. 1385), in ROBINSON, THE COMPLETE WORKS OF GEOFFREY CHAUCER 632 (1933):

> Somtyme the world was so stedfast and stable
> That mannes word was obligacioun.

[146] See, *e.g.*, UCC §§ 2-725(1) (short statutes of limitation); 9-318(4) (prohibition of account assignments not permitted). Indeed the exclusion of the buyer's family from his warranty protection *has* simply been forbidden by § 2-318.

[147] Compare the Code's express ducking of any express position in the developing area of privity of warranty (beyond protecting household members), UCC § 2-318, comment 3. It is not as if this warranty area is one without political sensitivity or strong feelings. See, *e.g.*, Condon, *The Practical Impact of the Proposed Uniform Commercial Code on Food Poisoning Cases*, 5 FOOD DRUG COSM. L.J. 213 (1950); Dierson, *Report on the Proposed Uniform Commercial Code*, 6 FOOD DRUG COSM. L.J. 943 (1951); Duesenberg, *The Manufacturer's Last Stand: The Disclaimer*, 20 BUS. LAW. 159 (1964).

If one concludes, however, as I do, that if there is one sales-contract provision to which 2-302 does not apply it is the warranty disclaimer, then both kinds of clauses dealt with in the official-comment cases are totally regulated by sections of the Code other than 2-302. But may one not reason by analogy from the warranty-disclaimer and remedy-limitation clauses at issue in the official comment cases to clauses which are "like" them but do not have any specific applicable section of the Code?[148] The answer, I think, is no. First let us recall that we are not talking here of procedural unconscionability. Assuming that a certain level of bargaining nastiness is reached, any harsh clause may be strikable; but we are talking, remember, of form contracts, or at most of contracts of adhesion, contracts whose provisions cannot be handled in any per se simplified manner. Thus we are speaking of what is "like" a warranty disclaimer as a substantive provision. What is that? Basically, it is a provision which shifts a risk from party A to party B when party A is, arguably, better able to appreciate, avoid and stand that risk.[149] Put briefly, can we assume that a provision is unconscionable and voidable if (a) it is in a form contract or a contract of adhesion and (b) it makes the poorer party stand a substantial loss which the richer party could stand better? Moreover, it must be recalled that the Code most specifically did not declare warranty disclaimers and remedy limitations void. Instead it regulated them in detail. Is one to assume that while the paradigms are to be regulated the clauses "like" them are to be voided instead? What analogy suggests here is not similarity of treatment, but unspecified variation instead.

The official-comment cases do illustrate, if nothing else, the responses of judges in the throes of one of the dilemmas of the judicial process, and that, if nothing else, is what they were designed to illustrate. In the first draft of 2-302,[150] there was no mention of illustrative cases, though Professor Llewellyn showed himself aware of what judges do when they face an apparent duty to reach harsh results in a particular case.[151] The very next time comments to a draft were prepared, however, they closed with the following paragraph:

---

[148] *Cf.* Braucher, *Sale of Goods in the Uniform Commercial Code*, 26 LA. L. REV. 192, 204 (1966):

> Perhaps these more specific provisions [§§ 1-102, 2-309, 2-616, 2-718, 2-719] help give content to the word unconscionable. Perhaps it can be held to contracts and clauses which have similar vices.

But he continues with some skepticism: "Certainly it is not a warrant for judicial price control. But people worry and you can see why."

[149] See Franklin, *supra* note 115, for an excellent particularized summary of what might be wrong with warranty disclaimers.

[150] MIMEO 1941 DRAFT § 1-C.

[151] MIMEO 1941 DRAFT § 1-C, comment A(7).

> Illustrations are needed, and they should indicate also the degree to which the courts have gone in avoiding the effect of forms which had been signed but which were felt by the courts to be unconscionable in the circumstances.[152]

Whatever else was then intended, therefore (and the "also" does imply a nonunitary purpose), illustration of the judges' pre-Code harshness-evading techniques was, from the beginning, one of the reasons for the projected cases' inclusion. Thus it was not surprising that while the very next version of the comments did not mention the point,[153] when they were violently revised the following year they began as follows:

> This section is intended to make it possible for the courts to police explicitly against the contracts or clauses which they find to be unconscionable. In the past such policing has been accomplished by adverse construction of language, by manipulation of the rules of offer and acceptance or by determinations that the clause is contrary to public policy or to the dominant purpose of the contract. This section is intended to allow the court to pass on the unconscionability of the contract or particular clause therein and to make a conclusion of law as to its unconscionability. . . . The underlying basis of this section is illustrated by results in cases such as the following.[154]

There then followed the citation of ten cases, each accompanied by its own individual brief descriptive tag—indeed *the* ten cases and tags, which, still pristinely unchanged, grace the present first comment to section 2-302.[155]

Assuming, therefore, that the ten cases are to illustrate what judges do when faced with appealing fact situations and unhelpful legal doc-

---

[152] AMERICAN LAW INSTITUTE, UNIFORM REVISED SALES ACT, Informal Appendix, at 12 (3d Draft 1943).

[153] See 1949 DRAFT § 2-302.

[154] 1950 DRAFT § 2-302, comment 1.

An interesting foreshadowing of this language may be found in 2 GENY, METHODE D'INTERPRETATION § 174, at 420 (2d ed. 1954). The language is not, however, quite close enough to support attributing it as the source of the comment.

[155] This fidelity to original reading has at least one untoward result: Kansas Flour Mills Co. v. Dirks, 100 Kan. 376, 164 Pac. 273 (1917), has remained misdescribed to date. Its tag says that under a clause permitting the buyer, upon the seller's default, to "extend time of delivery, cancel contract, or buy in for [seller's] account," 100 Kan. at 376-77, 164 Pac. at 274, "in a rising market the court permitted the buyer to measure his damages for non-delivery at the end of only one 30-day postponement." In fact, the buyer never attempted to extend the contract for more than one additional period, it was a fifteen-day period in any event, and the propriety of that extension was expressly sustained by the Kansas Supreme Court which indeed reversed the trial court's contrary holding on that particular point.

In addition, one would think that even in 1950 it was not necessary to use a collection of illustrative cases the most recent of which was decided in 1937 and over half of which were decided before 1929. One suspects the mining of a secondary source, but like an archaeologist finding a large selection of Egyptian shards in a Greek pit, one can suspect a prior collector without knowing who he was.

trines, how do they do that job? First, they illustrate only "adverse construction of language," which is only one of the four evasive techniques named in the comment.[156] They do not exemplify "manipulation of the rules of offer and acceptance," or any findings that a clause is contrary to "the dominant purpose of the contract." Certainly none of the cases says that the clause is bad as "contrary to public policy." Thus, the cases are apparently not designed to be exhaustive on the subject of manipulative techniques. What they do illustrate, however, and quite well,[157] is the skewing of legal doctrine that may be caused by an emotional pressure to get a more heartwarming particular result. It cannot be denied that uncertainty of a particularly virulent kind enters the picture when the basis of a decision and its stated basis part company.[158] This uncertainty, coupled with the distorting effect on legal doctrine of generous manipulations to get "good" results (that is, the "pore-ole-widder-lady" syndrome [159]) was fully appreciated by the chief draftsman before the drafting started,[160] was adverted to in the very first comment ever appended to foetal 2-302,[161] and continued to be firm in his thought well after the Code was completed.[162] It was this tendency which the express and open invalidating power given to the judges was designed to prevent. As Professor Llewellyn put it, "covert tools are never reliable tools." [163]

If, therefore, this uncertainty and skewing of doctrine could be prevented by something like section 2-302, its inclusion in the Code, despite the difficulties involved, might be justified. How much of a gain, however, is likely when there is substituted for the court's obligation to give false reasons for its behavior, a specific power to give no reason at all? An answer may come from what the courts thus far have done with their shiny new weapon in the very few cases in

156 UCC § 2-302, comment 1.

157 Cases showing the manipulation of the rules of offer and acceptance might have better illustrated the point. See, *e.g.*, Alexander Hamilton Institute v. Jones, 234 Ill. App. 444 (1924) (a save-the-correspondence-school-student case).

158 See Andrews Bros. v. Singer & Co., [1934] 1 K.B. 17 (C.A.), in which Scrutton, L.J., said of the warranty disclaimer in issue, "Clause 5 is, I take it, a sequel to *Wallis, Son & Wells v. Pratt & Haynes* [[1910] 2 K.B. 1003; [1911] A.C. 394]" and went on thereafter to say,

> Those advising the present defendants . . . appear to have thought that by the inclusion of the word "conditions" [as suggested by the *Wallis, Son & Wells* case] . . . liability would be excluded. . . . [1934] 1 K.B. at 21-22.

Scrutton found the advisors wrong, again by "interpretation of language."

159 See, for instance, its illustration for recent law students in FULLER & BRAUCHER, BASIC CONTRACT LAW 792 (1964), where Fox v. Grange, 261 Ill. 116, 103 N.E. 576 (1913) is printed (masquerading as a cancellation-of-waiver case).

160 See Llewellyn, Book Review, 52 HARV. L. REV. 700, 703 (1939).

161 See MIMEO 1941 DRAFT § 1-C, comment A(7).

162 See LLEWELLYN, THE COMMON LAW TRADITION 364-65 (1960) quoting at length from his Book Review, *supra* note 160.

163 *Ibid.*

which it has thus far figured. But first it will be useful to consider another place in the law where judges were, arguably, given the power to decide cases on the basis of the high level abstraction, "unconscionability."

UNCONSCIONABILITY IN EQUITY

It is the most common thing in the world for a commentator on section 2-302, apparently impelled by the obvious fact that the section itself embodies no noncircular standards, to lessen his nervousness by pointing to the equity court's old and well-established unconscionability doctrine as a sufficient illumination of the Code provision.[164] "After all," he seems to say, "why get excited? This is nothing new."[165] Moreover, among the works most frequently pointing with elaborate but unelaborated calm to the equity doctrine are substantially all of the "official" state studies, undertaken generally for the guidance of legislatures.[166] One gets the impression, in fact, that everyone who thought of mentioning the equity doctrine mentioned it.

---

[164] A fair sampling would include 1 HAWKLAND, A TRANSACTIONAL GUIDE TO THE UNIFORM COMMERCIAL CODE § 1.1603, at 46 (1964); Charney, *How to Make a Contract Under the U.C.C.*, 16 BROOKLYN BARRISTER 18, 27 (1964); Held & Torbert, *Comparison of Articles Two and Three of the Uniform Commercial Code With the Law of Alabama*, 7 ALA. L. REV. 271, 287 (1955); Kuenzel, *Uniform Commercial Code: Its Effect on Florida's Existing Law of Sales*, 36 FLA. B.J. 1020, 1027 (1962); Leflar, *The Commercial Code in Arkansas*, 14 ARK. L. REV. 302, 308 (1960); Lorensen, *The Uniform Commercial Code Sales Article Compared With West Virginia Law*, 64 W. VA. L. REV. 142, 143 (1962); Peters, *Remedies for Breach of Contracts Relating to the Sale of Goods Under the Uniform Commercial Code: A Roadmap for Article Two*, 73 YALE L.J. 199, 202 n.10 (1963); Stockton, *Sales Under the Uniform Commercial Code: Significant Changes*, 20 ALA. LAW. 352, 359 (1959); Weeks, *supra* note 139, at 517; Note, 58 DICK. L. REV. 161 (1954); Note, 22 TENN. L. REV. 776, 793 (1953); Project, 10 U.C.L.A.L. REV. 1087, 1131 (1963).

[165] See, *e.g.*, Charney, *supra* note 164, at 27 ("Actually all this section . . . ."); Leflar, *supra* note 164, at 308 ("This amounts essentially to . . . .").

This "nothing new" argument must be distinguished from the "nothing new" argument about the contractual manipulations of common-law judges to get just results. See, *e.g.*, § 2-302, comment 1; Hawkland, *Uniform Commercial "Code" Methodology*, 1962 U. ILL. L.F. 291, 305-07 (1962); Note, 109 U. PA. L. REV. 401, 402 (1961). This latter argument was broached in Llewellyn's first commentary on a proto-2-302, MIMEO 1941 DRAFT § 1-C, comment A(7).

[166] *E.g.*, *California Annotations to the Proposed Uniform Commercial Code*, in [CALIFORNIA] SENATE FACT FINDING COMMITTEE ON JUDICIARY, SIXTH PROGRESS REPORT TO THE LEGISLATURE 41-42 (1961); STARR, STUDY AND REPORT ON THE UNIFORM COMMERCIAL CODE 20, 71-72 (1959) (prepared for Connecticut Temporary Commission to Study and Report on the Uniform Commercial Code); ILLINOIS COMMISSION FOR THE UNIFORMITY OF LEGISLATION IN THE UNITED STATES, ILLINOIS ANNOTATIONS TO THE UNIFORM COMMERCIAL CODE 49 (1960); KENTUCKY LEGISLATIVE RESEARCH COMMISSION, UNIFORM COMMERCIAL CODE, ANALYSIS OF EFFECT ON EXISTING KENTUCKY LAW 43 (Research Publication No. 49, 1957); MASSACHUSETTS ANNOTATIONS TO THE PROPOSED UNIFORM COMMERCIAL CODE 31 (1953) ("Prepared by a Group of Massachusetts Lawyers and Law School Professors"); STEINHEIMER, MICHIGAN SALES LAW AND THE UNIFORM COMMERCIAL CODE 36 (1963); BAR ASSOCIATION OF NEW HAMPSHIRE, NEW HAMPSHIRE ANNOTATIONS TO THE PROPOSED UNIFORM COMMERCIAL CODE 24 (1959); NEW JERSEY UNIFORM COMMERCIAL CODE STUDY COMMISSION, SECOND REPORT TO THE GOVERNOR, THE SENATE AND THE ASSEMBLY OF THE STATE OF NEW JERSEY 52 (1960); HOGAN & PENNY, ANNOTATIONS OF THE UNIFORM COMMERCIAL CODE TO THE STATUTORY AND DECISIONAL LAW OF NEW YORK

This near unanimity of belief in the relevance of equity unconscionability is all the more striking in that neither section 2-302 nor its accompanying commentary makes any mention of it. Once upon a time it did, very clearly and explicitly, but the reference came late in the section's drafting history and didn't last very long. The May 1949 Draft's first accompanying comment began:

> This section is intended to apply to the field of Sales the equity courts' ancient policy of policing contracts for unconscionability or unreasonableness.[167]

That sentence lasted just about a year. In the very next draft of this comment, though the remainder of the paragraph in which it appeared remained wholly unchanged, the quoted sentence was deleted in its entirety.[168] And that sums up the entire history of overt references to the equity unconscionability doctrine in the Code;[169] it never appeared again.[170]

Well, not quite. If one looks at the current version of the Code, while there is nothing about the equity doctrine in section 2-302, and while the comments accompanying it make no such reference, and while none of the ten cases cited and described in the first official comment as "illustrating" the "underlying basis of the section" had anything to do with a request for specific performance or even came up in equity, there is one reference which may be significant. In the 1952 version of the official comments a key segment was changed from the way it had appeared in the 1950 version, as indicated below by italics:

STATE 35 (1961); OHIO LEGISLATIVE SERVICE COMMISSION, OHIO ANNOTATIONS TO THE UNIFORM COMMERCIAL CODE 17 (1960); OREGON STATE BAR COMMITTEE ON CONTINUING LEGAL EDUCATION, UNIFORM COMMERCIAL CODE HANDBOOK 35-36 (1963); PENNSYLVANIA GENERAL ASSEMBLY JOINT STATE COMMISSION, PENNSYLVANIA ANNOTATION TO THE PROPOSED UNIFORM COMMERCIAL CODE 18 (1953); FURLOW, REPORT OF THE SPECIAL COMMISSION [of the State of Rhode Island] TO STUDY AND REPORT ON THE UNIFORM COMMERCIAL CODE 22 (1960). Some state studies seemed to see no particular problem, *e.g.*, VIRGINIA CODE COMMISSION, REPORT ON UNIFORM COMMERCIAL CODE 19 (House Doc. No. 28, 1955). Others do see some definitional problem but nevertheless do not rely on the old equity practices. *E.g.*, MISSOURI GENERAL ASSEMBLY COMMITTEE ON LEGISLATIVE RESEARCH, PROPOSED UNIFORM COMMERCIAL CODE: ITS EFFECT UPON COGNATE MISSOURI STATUTES 48 (1954); TEXAS LEGISLATIVE COUNCIL, ANALYSIS OF ARTICLE 2 OF THE UNIFORM COMMERCIAL CODE 55-58 (1953).

[167] 1949 DRAFT § 2-302, comment 1. At no time in the drafting history of § 2-302 was the equity practice adverted to in the text of the statute itself.

[168] See 1950 DRAFT § 2-302, comment 2. (The whole comment had dropped down to second place.)

[169] References to the equity practice of "reformation" did appear in earlier drafts, beginning with 1943 DRAFT § 24, and were not eliminated until 1948 DRAFT § 23. But these were not references to the equity unconscionability doctrine.

[170] In fact, at the May 1951 meeting an effort was made to assure the section against any misinterpretation that solely an equity application was meant. See AMERICAN LAW INSTITUTE, UNIFORM COMMERCIAL CODE, MAY MEETING REVISIONS TO PROPOSED FINAL DRAFT No. 2, § 2-302 (1951):

> The words "refuse to enforce" are to be reconsidered for rephrasing to avoid inference that it deals only with the question of specific performance.

No change came out of this reconsideration, however.

> The principle is one of the prevention of *oppression and* unfair surprise (*Cf. Campbell Soup Co. v. Wentz, 172 F.2d 80, 3d Cir. 1948*) and not of disturbance of allocation of risks because of superior bargaining power.[171]

This, of course, is how this segment reads today.[172] Now, the *Campbell Soup* case *was* a case in equity, and in fact had precisely to do with a request for specific performance. And, moreover, in that case the Third Circuit [173] denied specific performance on the explicit ground that the contract involved in the case was "unconscionable." [174] "That equity does not enforce unconscionable bargains," said the court without elaborate citation,[175] "is too well established to require elaborate citation." [176]

Taking the identity of the words "unconscionable" in the Code section and in the equity doctrine, together with this rather obscure reference to the *Campbell Soup* case, there is some justification for the widespread belief that section 2-302 is just tried and true equity applied to the field of Sales. It does seem to me that if the draftsmen had meant to signal the importation into the statute of such a vast body of decisions and learned commentary as the equity doctrine involves [177] they might have chosen a reference less coy than a "*cf.*" citation [178] to a single equity case. But let us assume that the doctrine was meant to be applicable.

Now, as a rough guess I would say that there are as many cases dealing with denials of specific performance as stars in the heavens or sand by the sea. The divers reasons given for such refusals are almost as

---

[171] *Compare* 1952 DRAFT § 2-302, comment 1, *with* 1950 DRAFT § 2-302, comment 1.

[172] See UCC § 2-302, comment 1.

[173] In an opinion by Judge Goodrich, at that time Chairman of the Editorial Board and Director of the American Law Institute. See 1949 DRAFT, Foreword at v.

[174] The lower court had rested its decision upon the ground that the carrots involved in the case were not "unique" enough to justify specific performance. The Court of Appeals specifically rejected that ground for dismissing the bill, choosing instead to rely on the unconscionability doctrine. See Campbell Soup Co. v. Wentz, 172 F.2d 80, 82 (3d Cir. 1948).

[175] The citation (172 F.2d at 83 n.12) is limited to two treatises: 4 POMEROY, EQUITY JURISPRUDENCE § 1405a (5th ed. Symons 1941), and 5 WILLISTON, CONTRACTS § 1425 (rev. ed. 1937).

[176] 172 F.2d at 83.

[177] Treatments of various degrees of completeness, often accompanied by massive case citations, are found in 2 CHAFEE & SIMPSON, CASES ON EQUITY 1173-93, 1345-88 (1934); CLARK, EQUITY §§ 168-70 (1919); DE FUNIAK, MODERN EQUITY §§ 94, 95 (2d ed. 1956); FRY, SPECIFIC PERFORMANCE §§ 387-459 (6th ed. Northcote 1921); MCCLINTOCK, THE PRINCIPLES OF EQUITY §§ 69-72 (2d ed. 1948); 3 POMEROY, EQUITY JURISPRUDENCE §§ 926-28 (5th ed. Symons 1941); POMEROY, SPECIFIC PERFORMANCE OF CONTRACTS §§ 40, 46, 175-97 (3d ed. 1926); SNELL, PRINCIPLES OF EQUITY 501-06, 551-52 (25th ed. Megarry & Baker 1960); WALSH, EQUITY § 104 (1930); Annot., 65 A.L.R. 7 (1930).

[178] Which means, I assume, as it always has, something like "this fits here, but I can't tell how." *Cf.* HARVARD LAW REVIEW ASSOCIATION, A UNIFORM SYSTEM OF CITATION § 27:2:4 (10th ed. 1958).

extensive. To be enforced specifically a contract must first be a contract, and thus the issues of capacity, consideration, agreement and formality are as relevant to equitable as to legal contract actions.[179] Certainly those failings in the contracting behavior of one of the parties which would prevent relief being given him at law will ordinarily prevent his procuring specific performance in equity.[180] But in addition to these considerations which are applicable both to legal and equitable actions are others which are recognized only in courts of equity as applicable to specific performance. Pomeroy saw these additional considerations merely as applications of "the grand and far-reaching principle that he who seeks equity must do equity," [181] but whatever their genesis, they are various and numerous:

> the specific performance of a contract will be refused when the plaintiff has obtained the agreement by sharp and unscrupulous practices, by overreaching, by concealment of important facts, by trickery, by taking undue advantage of his position, or by other means which are unconscientious; and when the contract itself is unfair, one-sided, unconscionable, or affected by any other such inequitable feature, and where the specific enforcement would be oppressive or harsh upon the defendant, or would prevent the enjoyment of his own rights, or would in any other manner work injustice.[182]

Briefly put, when one examines any number of these equity cases [183] at all it becomes abundantly clear that over and above fraud, misrepresentation, mistake and duress there is a whole universe of *kinds of bargaining* which, while not sufficient to justify the voiding of a contract, will support a refusal specifically to enforce it, and that beyond the illegality and "against public policy" rubrics of law, are *kinds of contracts* which equity will not affirmatively aid.[184]

Within the ambit of those factors of contract-procuring behavior which would result in a denial of specific performance, a bewildering

[179] See POMEROY, SPECIFIC PERFORMANCE OF CONTRACTS §§ 51-161 (3d ed. 1926).

[180] See *id.* §§ 209-28 (misrepresentation), 229-66 (mistake), 267-79 (fraud), 280-87 (illegality).

[181] *Id.* § 40.

[182] *Ibid.* This, by the way, is merely a somewhat expanded version of the Pomeroy summation cited by the court in the *Campbell Soup* case. See note 175 *supra.*

[183] I have not come close to reading all of the thousands of cases dealing with unconscionability in equity. Moreover, those which I have read have not been selected according to any plan of reasoned randomness. A more extensive study of these cases might yield, therefore, further or other generalizations.

[184] The classic citation for this power to deny specific performance on grounds insufficient to justify cancellation of the contract is Day v. Newman, 2 Cox Ch. 77, 30 Eng. Rep. 36 (Ch. 1788), in which, faced with cross bills for specific performance and for cancellation, the Master of the Rolls dismissed both bills. (When the plaintiff refused to rescind the contract in exchange for an award of costs, the bills were dismissed without costs, too.) See for a more modern and equally clear example, Kleinberg v. Ratett, 252 N.Y. 236, 169 N.E. 289 (1929).

number of permutations work to inform the chancellor's discretion. In these cases one runs continually into the old,[185] the young,[186] the ignorant,[187] the necessitous,[188] the illiterate,[189] the improvident,[190] the drunken,[191] the naive [192] and the sick,[193] all on one side of the transaction, with the sharp and hard [194] on the other. Language of quasi-fraud [195] and quasi-duress [196] abounds. Certain whole classes of presumptive sillies like sailors [197] and heirs [198] and farmers [199] and

---

[185] *E.g.*, Blackwilder v. Loveless, 21 Ala. 371 (1852); Banaghan v. Malaney, 200 Mass. 46, 85 N.E. 839 (1908); Hemhauser v. Hemhauser, 110 N.J. Eq. 77, 158 Atl. 762 (N.J. Ch. 1932).

[186] *E.g.*, Clitherall v. Ogilvie, 1 Desaussure Ch. 250, 261 (S.C. 1792) ("Young heirs even when at age . . . .").

[187] *E.g.*, Fish v. Leser, 69 Ill. 394 (1873); Banaghan v. Malaney, 200 Mass. 46, 85 N.E. 839 (1908); Dysarz v. Janczarek, 238 Mich. 529, 213 N.W. 694 (1927); Fitzpatrick v. Dorland, 27 Hun (34 N.Y. Sup. Ct.) 291 (1882).

[188] *E.g.*, Blackwilder v. Loveless, 21 Ala. 371, 373 (1852) ("being a poor man . . . he [entered into the contract] to save his crop"); Fitzpatrick v. Dorland, *supra* note 187, at 292 ("pecuniarily embarrassed").

[189] *E.g.*, Dysarz v. Janczarek, 238 Mich. 529, 213 N.W. 694 (1927); Caton v. Wellershouse, 77 Pa. Super. 331 (1921).

[190] *E.g.*, Clitherall v. Ogilvie, 1 Desaussure Ch. 250 (S.C. 1792).

[191] *E.g.*, Knott v. Giles, 27 App. D.C. 581 (1906) ("habitual drunkard"); Moetzel & Muttera v. Koch, 122 Iowa 196, 97 N.W. 1079 (1904); see also Campbell v. Spencer, 2 Binn. (11 Pa.) 129, 133 (1809) (drunkenness not proved but "bargaining . . . amidst the drinking of bitters early in the morning").

[192] *E.g.*, Bartley v. Lindabury, 89 N.J. Eq. 8, 10, 104 Atl. 333, 334 (Ch. 1918) (farmer "unfamiliar with business methods"); Smedes v. Wild, 1 Livingston's Law Mag. 155 (N.Y. Sup. Ct. 1852) ("contract between a businessman and an inexperienced woman"); Campbell v. Spencer, *supra* note 191, at 133 ("and I do not like a contract by which a *farmer* is involved in the folly of buying a store of goods.").

[193] *E.g.*, Fitzpatrick v. Dorland, 27 Hun. (34 N.Y. Sup. Ct.) 291, 292 (1882) ("invalid, very infirm").

[194] *E.g.*, Banaghan v. Malaney, 200 Mass. 46, 85 N.E. 839 (1908); Kelley v. York Cliffs Improvement Co., 94 Me. 374, 47 Atl. 898 (1900) (even though defendant was a corporation); Wilson v. Bergmann, 112 Neb. 145, 198 N.W. 671 (1924); Hemhauser v. Hemhauser, 110 N.J. Eq. 77, 158 Atl. 762 (Ch. 1932) (stepchild of old "widder" lady); Falcke v. Gray, 4 Drew. 651, 62 Eng. Rep. 250 (Ch. 1859) (antique dealer, lady and valuable china). Of course, even if a court makes no special point of the plaintiff's behavior in any given case, getting a *very* good deal from the old, the sick, the drunk, etc., is itself hardly a character reference.

[195] *E.g.*, Pope Mfg. Co. v. Gormully, 144 U.S. 224, 237 (1892) ("the contract 'was an artfully contrived snare'"); Gabrielson v. Hogan, 298 Fed. 722, 725 (8th Cir. 1924) ("the Hogans were strangers").

[196] Pindall v. Waterman, 84 Ark. 575, 106 S.W. 964 (1907) (conveyance to attorneys with lynch mob in offing—set aside); Blackwilder v. Loveless, 21 Ala. 371, 373 (1852) ("to save his crop").

[197] *E.g.*, How v. Weldon, 2 Ves. Sr. 516, 518, 28 Eng. Rep. 330, 331 (Ch. 1754) ("a race of men loose and unthinking"). Compare the treatment of the same class in a different context, American maintenance-and-cure decisions, *e.g.*, Koistinen v. American Export Lines, Inc., 194 Misc. 942, 83 N.Y.S.2d 297 (N.Y. City Ct. 1948) (leaping from brothel window part of normal sailor's normal duties—under the circumstances). See generally GILMORE & BLACK, ADMIRALTY § 6-8 (1957).

[198] *E.g.*, Earl of Chesterfield v. Janssen, 2 Ves. Sr. 125, 28 Eng. Rep. 82 (Ch. 1750); Clitherall v. Ogilvie, 1 Desaussure Ch. 250, 261 (S.C. 1792) ("Young heirs even when at age are under the care of this court").

[199] *E.g.*, Koch v. Streuter, 232 Ill. 594, 83 N.E. 1072 (1908); Bartley v. Lindabury, 89 N.J. Eq. 8, 104 Atl. 333 (1918); Campbell v. Spencer, 2 Binn. (11 Pa.) 129 (1809).

women[200] continually wander on and off stage. Those not certifiably crazy, but nonetheless pretty peculiar,[201] are often to be found. And in most of the cases, of course, several of these factors appear in combination.[202] It might be assumed, therefore, that one setting out to find a body of decisions which might give contour and limits to a word like "unconscionability," at least insofar as that word might have something to do with the insufficiency of the bargaining process, would find in these cases riches beyond the dreams of judicial avarice. There is, however, one weakness in using these cases as a guide to the meaning of unconscionability in section 2-302: they are all irrelevant—for two reasons. First, the equity cases are of interest, if at all, *only* for giving outline to the limits of procedural unconscionability; they cannot define what kind of clause might be substantively unconscionable because they all involve only one form of substantive unconscionability—overall imbalance. Second, on procedural unconscionability, the dramatic situations which have produced the contracts which have produced the equity cases are exceedingly unlikely to be reproduced in a Sales context except on the very rarest of occasions, and thus their details do not inform the sales-contract decision a bit.

Procedural Unconscionability in Equity

It is a commonplace, even in the very best of circles,[203] to view with more than equanimity the application of equitable doctrines to actions at law. This is undoubtedly to some extent the natural by-product of the merger of legal and equitable procedures in modern codes,[204] but the trend can hardly be considered merely an offshoot of adjective reform. There are arguments, occasionally quite passionate,[205] that the importation has not gone far enough,[206] and it is very clear that the extended use of at least some equitable doctrine is becoming more

200 Smedes v. Wild, 1 Livingston's Law Mag. 155 (N.Y. Sup. Ct. 1852); Friend v. Lamb, 152 Pa. 529, 536, 25 Atl. 577, 579 (1893) (married woman with only "very recent emancipation").

201 *E.g.*, Wilson v. Bergmann, 112 Neb. 145, 198 N.W. 671 (1924); Miller v. Tjexhus, 20 S.D. 12, 104 N.W. 519 (1905). *Cf.* Green, *Proof of Mental Incompetency and the Unexpressed Major Premise*, 53 YALE L.J. 271, 306-07 (1944), arguing that when courts say they have found "mental incompetency" they usually mean only that they have found a peculiar deal.

202 See, *e.g.*, Banaghan v. Malaney, 200 Mass. 46, 85 N.E. 839 (1908), involving an "aged, inexperienced and ignorant woman."

203 See, *e.g.*, FULLER & BRAUCHER, BASIC CONTRACT LAW 754 (1964).

204 *E.g.*, FED. R. CIV. P. 2; MO. R. CIV. P. 42.01.

205 *E.g.*, Newman, *The Place and Function of Pure Equity in the Structure of Law*, 16 HASTINGS L.J. 401 (1965).

206 See NEWMAN, EQUITY AND LAW: A COMPARATIVE STUDY 115 (1961); Puig Brutau, *Juridical Evolution and Equity*, in ESSAYS IN JURISPRUDENCE IN HONOR OF ROSCOE POUND 82 (1962).

frequent.[207] Insofar as the wisdom of this trend, or at least of its accelerated development, is questioned, it is generally on the ground that too much faith in increased judicial discretion (considered a hallmark of equity jurisprudence) as a simplified way to "justice" is dangerous, or at least over-sanguine.[208] On the other hand, some doctrines developed in courts of equity are perfectly applicable to law actions, and their exclusion therefrom is absurd. In other words, there is nothing in an "equitable" doctrine as such that particularly makes it unfitted for importation into an action which would historically have been an action "at law."

Merely that there is no a priori reason why doctrines developed in equity might not fit equally well in law actions, does not justify the jumped conclusion that all equitable doctrines fit equally well at law. Put more concretely for present purposes, the practice of denying specific performance in equity to contracts because of their "unconscionability," does not *necessarily* make any sense when applied to the law of Sales. It might be sensible. In fact, it isn't.

Almost without exception, actions for specific performance were (and are) brought with respect to transactions involving real property.[209] Article 2 of the Code governs "goods" only, and real property is not a species of "goods."[210] One well might argue that the subject

---

[207] See M.M.W., *Equity in the Commercial World*, 105 L.J. 627 (1955); Van Hecke, *Changing Emphases in Specific Performance*, 40 N.C.L. REV. 1 (1961); Note, 49 IOWA L. REV. 1290 (1964).

[208] See CARDOZO, NATURE OF THE JUDICIAL PROCESS 136 ("benevolent despotism" of Judges), 141 ("the judge . . . is not a knight-errant, roaming at will in pursuit of his own ideal of beauty or of goodness"); EVERSHED, ASPECTS OF ENGLISH EQUITY 16-17 (1954); LUNDSTEDT, LAW AND JUSTICE 30-39 (1952); Berolzheimer, *The Perils of Emotionalism*, in SCIENCE OF LEGAL METHOD 166, 185 (1921); Cohen, *Jerome Frank*, in COHEN, LAW AND THE SOCIAL ORDER 357, 362 (1931) ("uncontrolled discretion of judges would make modern complex life unbearable"); Wright, *Opposition of Law to Business Usages*, 26 COLUM. L. REV. 917, 917 n.1 (1926); Mann, Book Review, 80 L.Q. REV. 589, 590 (1964) ("the present unfortunate tendency towards a system of Cadi jurisprudence"). See also Cohn, *Frustration of Contract in German Law*, 28 J. COMP. LEG. & INT'L L. (3d ser.) 15, 23 (1946); Liberman, *Opportunity and Challenge to Bring Commercial Law in Step With Present Day Needs*, 62 COM. L.J. 221, 226 (1957), citing HEDEMANN, DIE FLUCHT IN DIE GENERALKLAUSEN [The Flight into the General Clauses] 1-4, 6-12, 46-52 (1933), on the temptations and dangers of broad discretionary standards in the hands of a burgeoning totalitarian state. The Hedemann book is in German only, which I can not read, so I have not read it. In this same connection, see the charming understatement in PRAUSNITZ, THE STANDARDIZATION OF COMMERCIAL CONTRACTS IN ENGLISH AND CONTINENTAL LAW 6-7 (1937): "In 1934 General Goering . . . said: 'The law and the will of the Fuehrer are one.' This maxim by itself may influence the interpretation of certain contracts."

It should not be thought that the above-cited works are necessarily simple-mindedly against judicial discretion; in fact portions cited are more often than not caveats tacked onto the explicit recognition of the need for judicial discretion, *e.g.*, CARDOZO, *op. cit. supra* at 124, 129, 136-58.

[209] McCLINTOCK, THE PRINCIPLES OF EQUITY § 44 (2d ed. 1948); POMEROY, SPECIFIC PERFORMANCE OF CONTRACTS § 10 (3d ed. 1926). *But see* Van Hecke, *supra* note 207; Campbell Soup Co. v. Wentz, 172 F.2d 80 (3d Cir. 1948).

[210] UCC §§ 2-102, 2-105.

matter of a transaction might reasonably have an effect upon the form it takes, and upon the legal rules which might develop to confine, define and delimit that form. Put into contract-teacher terminology, widgits and Blackacre are not the same, are not dealt with by parties in the same way and (at least arguably) ought not to be treated identically in law.[211] In still other words, the successful struggle to unhorse Sales which the Code represents[212] may have resulted in its unhappy implantation in alien soil.

Land transactions and chattel transactions are different because land and chattels are different. This is not to say that they are in all ways different, or that they can never reasonably be treated by the law as if they were not different, but only to suggest that the realty-personalty dichotomy is not arbitrary. To the extent that real distinctions do exist between the two subject matters, the learning surrounding the equity unconscionability doctrine, a specific performance doctrine and a land doctrine, may be inapplicable, and if applied to Sales, misleading. This depends, of course, on the extent to which the two subject matters do differ.

First, as a general rule a real estate transaction is likely to be economically significant for at least one, and often both, of the parties to it. It would be my guess that both at the time the equity doctrine was developing, and today, the largest single transaction which most people will enter into during their lives will be a real estate transaction of one sort or another. It is true that the purchase of the ubiquitous automobile is no trivial transaction today, but more is ordinarily involved in the family's home purchase. Even between professional real estate traders, each transaction is likely to involve more dollars than in most other businesses where the units of "merchandise" are smaller. Thus a disparity between "value" and price would more likely be a serious economic hardship with respect to land than elsewhere.

Second, as a general rule land transactions are more likely to be one-in-a-lifetime transactions for at least one of the parties than the commercial transactions the Code is primarily designed to govern. This would tend to limit those protections against overreaching which follow from a businessman's desire to build a following, to establish and maintain as continual and continuous a relationship as possible. In

211 This suggestion that the subject matter of a transaction determines the law which will grow up around it, and that such law may be absurd when applied to a totally different kind of transaction is, of course, hardly original. A most completely developed exploration of the insightful point that sales of widgitis differ from dickers over Dobbin is to be found in Llewellyn, *The First Struggle to Unhorse Sales*, 52 HARV. L. REV. 873 (1939).

212 That business had progressed beyond Dobbin was appreciated by some rather early in that progression. See Note, 27 COLUM. L. REV. 430, 435 n.20 (1927) ("disappearance of the horse-trade manner of doing business").

other words, businessmen's accommodations, even when dealing with consumers, are lessened when the transaction, because of the nature of its subject matter, is relatively unlikely to be repeated between the same parties.[213]

Third, in spite of the homogenization of land in the modern world, and its transformation into more and more of a commodity (like the Code's "goods"), land is not just *treated* as unique in equity,[214] it ordinarily *is* unique. That is, if the character of an object is dependent upon its surroundings, and both it and its surroundings are immovable, then no other object is the same, or can be. One can of course buy land for purposes with respect to which both its inherent character and its position is relatively irrelevant. I would be willing to guess that much Iowa corn land, as well as much Texas grazing land, is substantially interchangeable (or at least so it seems to one raised in New York) and it is very hard to see the distinction between various addresses in Levittown. But it often is the case that land is chosen for either its intrinsic character (*e.g.*, soil richness) or its relational character (*e.g.*, nearness to a particular school). Because these factors are effectively irrevocable and irreparable (*i.e.*, it is hard to turn sand to loam or move a school), transactions with respect to this type of commodity are transactions with respect to a semi-permanent personal commitment of some sort, and the legal rules that govern it would reasonably tend to be hedged with additional restrictions.[215]

Most important, real property is likely to be the only thing that relatively unsophisticated people have which is worth tricking them out of. Farmers have farms and old ladies have old homesteads. The equity cases are replete with factual patterns involving the old being bilked,[216] and farmers sweet-talked into ruinous trades.[217] Courts would be most solicitous to impede land transfers by the poor sillies of the world.[218]

---

[213] *Cf.* Jones, *Merchants, The Law Merchant, and Recent Missouri Sales Cases: Some Reflections*, 1956 WASH. U.L.Q. 397, 411-18; Jones, *Three Centuries of Commercial Arbitration in New York: A Brief Survey*, 1956 WASH. U.L.Q. 193, 218-19. See also FRIEDMAN, CONTRACT LAW IN AMERICA 46-48 (1965), on the effect of nonrepetitiveness on the morals of Wisconsin real estate brokers.

[214] McCLINTOCK, *op. cit. supra* note 209; POMEROY, *op. cit. supra* note 209.

[215] This discussion does not take into account the exceedingly powerful sentimental-mystical aspects of land, the Tara and Mother Russia complexes of *Gone With the Wind* and *War and Peace*, for instance. These feelings, moreover, are apparently not even in any obvious relation to the objective charm of the land involved. See ROLVAAG, GIANTS IN THE EARTH (1927) (Dakota prairie). See also FRIEDMAN, *op. cit. supra* note 213, at 35, on the transformation of land from commodity to "differentiated space value" in Wisconsin.

[216] See, *e.g.*, cases cited note 185 *supra*.

[217] See, *e.g.*, cases cited note 199 *supra*.

[218] And it is a "commodity" which often needs a judicial imprimatur of some sort to render it resalable, whence actions to quiet title which apply only to an "estate or interest in real property, whether the same be legal or equitable." Mo. ANN. STAT. § 527.150 (1953).

It is out of these special attributes of land, making up the *Gestalt* of real property (as opposed to the "goods" of the Code), that there arise those repeated dramatic vignettes with which the Chancellors were continually faced—the abused old and unsophisticated young, the slicker and the farmer, the money lender and the expectant heir.[219] This cast of characters, to a large extent determined by the nature of the commodity, led to the various forms of overreaching which, while not quite adding up to fraud or duress, formed the pictures of bargaining processes which the chancellors declared "unconscionable." But mark: *all of these are pictures of individual overreachings.* In other words, more important than the uniqueness of each piece of land (but connected with it) is *the uniqueness of each land transaction.* The dramatic situations which were presented and decided under the equity unconscionability doctrine were most particularly those kinds of overreaching which take place, and can only take place, when there *is* individualized bargaining. The equity criteria are fitted only to nonmass transactions.

And that is precisely what the Code in general and section 2-302 in particular is *not* designed to cover. The unconscionability section of the Code is primarily focused on the merchant-to-merchant form-pad deal, the merchant-to-consumer adhesion transaction, the modern mass-sale transaction.[220] To decide whether one of these mass transactions is to be allowed to stand, the discriminations and discussions by the equity courts of various gradations of quasi-fraud and quasi-duress are about as useful as a goiter. Section 2-302 is a child of the mass transaction, and the state of health of little old ladies and the shade of rapaciousness of their favorite nephews is not going to inform one's decision. Thus, all of the jolly references to the good old equity doctrine,[221] if they are supposed to indicate a source for determining procedural unconscionability under the Code, are woefully misguided and misguiding. Equity dealt with the pathology of bargaining. The Code deals with the pathology of nonbargaining.

Substantive Unconscionability in Equity

If, then, the references to the equity doctrine are to be other than delusive, the mass of equity cases must help to define the kinds of con-

[219] See notes 185-202 *supra.*

[220] *Cf.* Latty, *Sales and Title and the Proposed Code,* 16 LAW & CONTEMP. PROB. 3, 19 n.78 (1951); Mentschikoff, *Highlights of the Uniform Commercial Code,* 27 MODERN L. REV. 167, 171 (1964); Pound, *The Role of the Will in Law,* 68 HARV. L. REV. 1 (1954). It is of some interest, I think, that the only thing left to a consumer these days which is anything like a dicker over Dobbin is a bicker over Blackacre.

[221] See note 167 *supra.*

tracts and contract clauses (as distinguished from the kinds of contracting behavior) which are unconscionable. Alas, that hope is also bootless. There is only one thing which equity recognized as substantive unconscionability: inadequate consideration (or, to put it another way, "gross overall imbalance"). It is instructive, I think, that the *Campbell Soup* case,[222] the only link (other than mere verbal similarity) between section 2-302 and equity unconscionability is itself an extraordinarily striking case of overall imbalance. In that case the contract *was* totally one-sided.[223] But the hardship to the defendant farmer was in no way the result of any harshness in the contract, but solely the result of the fact that the market value of the commodity he had sold for future delivery had tripled by the time delivery was due. The soup company reserved the power to do all sorts of nasty things to farmer Wentz,[224] but it didn't try. The term that hurt him was the price term, the ony one, that is, which was presumably negotiable and fair when set.[225] In other words, even though there was no causal connection between the terms of the contract and the hardship on the defendant, the court nevertheless refused enforcement because the contract itslf was too one-sided. Thus *Campbell Soup* is not only typical of the equity cases in general in that the substantive vice in the contract is gross overall lopsidedness, but it is, so to speak, super-typical (one is tempted to say archetypal) in that the one-sidedness complained of was even irrelevant to the harshness complained of.[226]

This important fact, that all of the equity unconscionability decisions really depend upon a finding of inadequate overall consideration, has been obscured by the fact that the really live issue in this area, the subject of a controversy lasting centuries, was not whether inadequate consideration was a *necessary* cause of the denial of specific performance, but whether it was a *sufficient* cause.[227] No one doubted

---

[222] Campbell Soup Co. v. Wentz, 172 F.2d 80 (3d Cir. 1948).

[223] See the court's description of the allegedly objectionable contract terms. *Id.* at 83. Through the courtesy of counsel for Mr. Wentz, I was given a photostat of the entire original contract. As one would have guessed, the court picked out the "worst" provisions to quote, but the rest of the contract is hardly filled with favors to the farmer.

[224] The one which seems most to have impressed (or depressed) Judge Goodrich was a provision that any time Campbell for one reason or another could not take carrots, the farmer could not, without its consent, sell them elsewhere. *Id.* at 83 n.11.

[225] See *id.* at 81.

[226] Of course, one could also just say that the case was silly.

[227] See 1 AMES, CASES IN EQUITY JURISDICTION 264 n.2 (1904); CHAFEE & SIMPSON, *op. cit. supra* note 177, at 1173-93; DE FUNIAK, *op. cit. supra* note 177, § 95, at 222-23; FRY, *op. cit. supra* note 177, §§ 438-59; HANBURY, MODERN EQUITY 550-51 (8th ed. 1962); MCCLINTOCK, *op. cit. supra* note 177, § 71; MAITLAND, EQUITY 246-47 (1909); 3 POMEROY, EQUITY JURISPRUDENCE §§ 925-28 (5th ed. Symons 1941); POMEROY, SPECIFIC PERFORMANCE OF CONTRACTS §§ 192-97 (3d ed. 1926); SNELL, PRINCIPLES OF EQUITY 551-52 (25th ed. Megarry & Baker 1960); 1 STORY, EQUITY

that it was a necessary cause, and in the nature of the way in which the question was presented—as opposition to the specific performance of a contract—it would be rare that the unfairness of the exchange would not be at least implicitly asserted. Why else would the transaction be opposed? More important, if the assumption of the uniqueness of realty were taken seriously, the very opposition to the decree would be a testament to the defendant's decision that the transaction was unfair, and that subjective belief would, in equity as to land, be conclusive.[228] Thus the only factor of substantive unconscionability which could be presented in an action for specific performance was that of disproportion of price, *i.e.*, overall imbalance.[229] The Code draftsmen, however, quite specifically determined, after an early impulse to the contrary, that section 2-302 would be applicable not only to contracts which were unbalanced in an overall sense, but also to those containing single "unconscionable" clauses. Since under this approach a separate substantive determination must be made on a clause-by-clause basis, the equity doctrine's weighing technique is generally irrelevant.

To summarize, there are two separate social policies which are embodied in the equity unconscionability doctrine. The first is that bargaining naughtiness, once it reaches a certain level, ought to avail the practitioner naught. The second is directed not against bargaining conduct (except insofar as certain results often are strong evidence of certain conduct otherwise unproved) but against results, and embodies the doctrine (also present in *laesio enormis* statutes) [230] that the infliction of serious hardship demands special justification. The first of these social policies cannot be reflected in section 2-302 in any helpful way unless one takes the position that everything in a form contract or an adhesion contract is to be stricken upon the nondrafting party's request, for that is the type of transaction with which the section is designed to deal. The second policy, that harsh results not be permitted irrespective of the fairness of the bargaining process or the unfairness of the provision at the time of the drafting, is an attractive

JURISPRUDENCE §§ 244-49, 333 (13th ed. Bigelow 1886); WALSH, *op. cit. supra* note 177, § 104, at 482-89; Note, 25 VA. L. REV. 834 (1939); Annot., 65 A.L.R. 7, 80-97 (1930). One may also consult the exceedingly learned opinion of Chancellor Kent in Seymour v. Delancey, 6 Johns. Ch. 222 (N.Y. Ct. Ch. 1822), *rev'd*, 3 Cowen 445 (N.Y. Ct. Err. 1824), for a summary of the controversy up to that date.

228 *Cf.* MAITLAND, *op. cit. supra* note 227, at 238, suggesting that one may want particular realty even if the money offered in exchange, objectively considered, is adequate.

229 In fact, so strong was this imbalance element that specific performance might not be granted even if the contract were fair when made, if subsequent developments made it oppressive in operation. See, *e.g.*, Willard v. Tayloe, 75 U.S. (8 Wall.) 557 (1869) (decree granted, but with conditions); McCarty v. Kyle, 44 Tenn. 288 (1867).

230 See BUCKLAND, TEXTBOOK OF ROMAN LAW 486 (3d ed. Stein 1963).

one because of the ease of its administration; it is not at all hard to identify a harsh result when it has come about. The difficulty with adapting that doctrine to the Code provision is that substantially all of the important provisions in a normal sales contract are *potentially* exceedingly harsh. Generally they are inserted to determine who will stand a loss, perhaps a total loss, if a particular happening happens, or at least to give a huge litigation advantage to one of the parties should the question come up. Something as innocuous as a choice-of-law provision in a contract will operate harshly if the law chosen is unfavorable to one of the parties.[231] The same harshness is even more clearly inherent in a warranty disclaimer; if the warranty question becomes material and the disclaimer is upheld, the seller will win and the buyer will lose.[232] Thus "unconscionability" cannot be equated with "harshness" as an abstract matter. Certain particular clauses may indeed be declared impermissible as a matter of policy; that is how a usury statute operates, and consumer protection statutes embody numerous interdictions of specific contractual provisions.[233] But the hallmark of unconscionability cannot be the harshness of the result without more, because sales clauses are designed to be harsh. Unless one says that *all* losses should be split or spread [234] (as has been suggested in special contexts),[235] a harsh result without more, even if the result of an adhesion or form-contract provision, cannot identify the impermissible.[236]

231 See, *e.g.*, Siegelman v. Cunard White Star Ltd., 221 F.2d 189 (2d Cir. 1955).

232 A recent example is Williams v. Chrysler Corp., 148 W. Va. 655, 137 S.E.2d 225 (1964).

233 See CURRAN, TRENDS IN CONSUMER CREDIT LEGISLATION 312-22 (1965), a chart showing the sundry provisions barred in various states from retail installment sales contracts.

234 Or at least should be if the person hurt were "poor" perhaps. With respect to the apparent political difficulties of the open avowal of such a position, note the Code's propagandists' horrified response to any suggestion that the Code represented "class legislation." See, *e.g.*, Barney, *The Uniform Commercial Code*, 7 PORTLAND U.L. REV. 9, 10 (1961); Beers, *The New Commercial Code*, 2 BUS. LAW. 14, 17 (1947); Malcolm, *The Uniform Commercial Code*, 39 ORE. L. REV. 318, 322-23 (1960). See also note 11 *supra* on the allied "who-really-drafted-the-Code" controversy.

235 See, *e.g.*, Calabresi, *The Decision for Accidents: An Approach to Nonfault Allocation of Costs*, 78 HARV. L. REV. 713 (1965). This area also seems to have its emotional aspects. See Calabresi, *Fault, Accident and the Wonderful World of Blum and Kalven*, 75 YALE L.J. 216 (1965).

236 This simplistic resultant-harshness test was firmly rejected by the draftsmen of the Code. In the present revision of the Code, no clause may be stricken unless it was unconscionable "at the time it was made." UCC §2-302(1). While most likely implicit from the beginning, this was explicitly stated for the first time in 1955, very late in the drafting history. See AMERICAN LAW INSTITUTE, SUPPLEMENT NO. 1 TO THE 1952 OFFICIAL DRAFT OF TEXT AND COMMENTS OF THE UNIFORM COMMERCIAL CODE (1955) (containing the changes "voted by Enlarged Editorial Board [of the Institute] Oct. 29-31, Nov. 13-14, 1954"). The reason, the revisers said, was to "make it clear that . . . the court in making such a decision is not to apply hindsight but is rather to consider the question of unconscionability as of the date of formation of the contract." *Id.* at 8. Thus the doctrine of at least some of the equity cases, that subsequently occurring hardship alone might prevent a contract's specific enforce-

If the unconscionability of a clause at the time it was made cannot be determined by looking at its eventual harsh effect then the test of unconscionability to be applied to any individual clause of a commercial contract is no further clarified. When is a warranty disclaimer "unconscionable"? Not, obviously, when it succeeds in disclaiming a warranty, but when it is as a matter of social policy "bad" that the warranty be disclaimable. When is that? One can argue about the answer, but at least when the question is asked in that way, one is arguably arguing about the right sub-questions, not about the content of an *n*th level abstraction like "unconscionability." Alas, 2-302 steers the latter course.[237]

SECTION 2-302(2)

As we have seen, when the question is presented as a decision as to the "unconscionability" of a single contractual provision, the vacuousness of the standard is apparent. This led, eventually, to at least one attempt to modify the section to supply an internal method by which the definitional void might be filled. This, of course, was 2-302(2), and the very limited effect of this subsection helps to clarify even more, I think, the fundamental drafting misconception of section 2-302. Section 2-302(2) reads as follows:

> When it is claimed or appears to the court that the contract or any clause thereof may be unconscionable the parties shall be afforded a reasonable opportunity to present evidence as to its commercial setting, purpose and effect to aid the court in making the determination.

Its genesis is clearer than that of just about any other element in section 2-302. It did not appear in the 1950 draft of the Code. Late in January 1951, the Enlarged Editorial Board for the Code met before the Section of Corporation, Banking and Business Law of the American

---

ment, was rejected. And it was rejected even though the New York Law Revision Commission specifically suggested that after-the-fact harshness be covered by § 2-302. See ASSOCIATION OF THE BAR OF THE CITY OF NEW YORK, REPORT OF THE COMMITTEE ON THE UNIFORM COMMERCIAL CODE 27-28 (1962).

[237] One other difference between the equity doctrine and § 2-302 should be mentioned here. Under the equity doctrine, the result of a refusal to enforce was, at least theoretically, not total failure of the plaintiff's cause, but only a remission to his rights "at law." Such "right" at law in fact might not exist. One empirical study (dealing, however, with only fifty-six cases) has suggested that as a general rule one who loses in equity loses for good. Frank & Endicott, *Defenses in Equity and "Legal Rights,"* 14 LA. L. REV. 380 (1954). One suspects, however, that the Chancellors thought there was a real remedy at law, and that the litigants did too; else the actions for cancellation and the judges' agonizing over them make little sense. See, *e.g.*, Day v. Newman, 2 Cox 77, 83, 30 Eng. Rep. 36, 38 (Ch. 1788).

Bar Association for hearings.[238] On January 28, Walter D. Malcolm reported that the Section's Committee on the Code had just defeated a proposal to strike section 2-302 altogether, but only by a five-to-five vote.[239] Professor Robert Braucher of Harvard then rose.

> I have an additional suggestion . . . and was directed to present this by the Council of the Bar Association Section. That would be to add a second sentence to this provision, the purpose of which would be to try to help a court which passes on the question of whether a contract or a clause is unconscionable.
>
> To understand the setting in which it is working, the sentence which I would propose would be the second sentence in 2-302:
>
> > When it is claimed or appears to the court that the contract or any clause thereof may be unconscionable, the court may afford the parties an opportunity to present evidence as to its commercial setting, purpose, and effect as used.
>
> I think that violates Mr. Luther's principle that you should not have procedure in this Code, but if you are going to give the court a charter to inquire into whether the grey goods trade has regulated itself properly under the Worth Street Rules, or whether the form of contract used generally by the canners is unconscionable, it would be desirable to have some reminder that there are complications known in the trade, and that what appears on its face to be unfair or unconscionable may not be in the light of conditions in the trade.[240]

Professor Llewellyn's reaction to this suggestion was more than receptive; it would not be unfair to call it ecstatic:

> The Drafting Staff will welcome that, will welcome such a subsection. It clarifies definitely the meaning of the Section and addresses the court's attention to vitally important stuff.[241]

This reaction ought not to have been unanticipated. In his personal comments to the very first version of 2-302, Professor Llewellyn wrote:

---

[238] A transcript of these hearings exists in mimeographed form. Proceedings of the Larger Editorial Board of the American Law Institute, January 27-28, 1951 [hereinafter cited as Proceedings]. In addition, a report on this meeting was included at 6 BUS. LAW. 164 (1951), which included some quotations from the verbatim transcript. I shall give parallel citations to both sources.

[239] Proceedings at 171; 6 BUS. LAW. at 184.

[240] Proceedings at 173-74; 6 BUS. LAW. at 185.

[241] Proceedings at 174; 6 BUS. LAW. at 185. Note particularly Professor Llewellyn's use of the exceedingly nonpictorial "stuff."

> *The one-sided group of clauses which are fair, but are needed to give protection against bad law.*
>
> Many groups of clauses in very frequent use in the Sales field are utterly one-sided, but are, for all that, entirely fair because they correct in a reasonable way an unfortunate condition of the law. The most frequent of these are seller's clauses protecting against various types of business impossibility, and protecting against the obligation of delivery on credit to a buyer who has become a risk.[242]

At any rate, the draftsmen of section 2-302(2) seem to have felt that it filled a need.[243] It was no sooner proposed than integrated into the act, in almost exactly the words proposed by Professor Braucher,[244] and thereafter it was carried forward into the present draft of the Code without substantial change.

Certainly the idea was sound enough. If judges were to be given the power to regulate the agreements within industries on an *ad hoc* basis, then, as Professor Braucher suggested, it would be useful if they were given the opportunity to learn, if only on an *ad hoc* basis, a little something about the industries they were regulating. Obviously what the sponsors of this new subsection had in mind was testimony on the technical business requirements of particular business complexes, perhaps on the order of a statistical defense on the basis of long-time experience of, let us say, the very tough seller-oriented insecurity provision provided in the Worth Street Rules mentioned by Professor Braucher.[245] And to this extent section 2-302(2) serves an important purpose: it makes possible the resuscitation of a provision which, though to the uninitiated might appear unreasonable, has a particular, reasonable job to do in a particular industry.[246]

Does section 2-302(2), however, really give any help in defining substantive unconscionability in any case in which the question is likely

---

[242] MIMEO 1941 DRAFT § 1-C, comment A(6).

[243] Professor Braucher indicates that one of the needs it filled was to lessen dissatisfaction over the result in Campbell Soup Co. v. Wentz, 172 F.2d 80 (3d Cir. 1948), in which the unconscionability point had been decided without the help of much warning, briefing or evidence. Braucher, *Sale of Goods in the Uniform Commercial Code*, 26 LA. L. REV. 192, 203-04 (1966).

[244] The only change was that the last two words "as used" were replaced by the phrase, "to aid the court in making the determination." See AMERICAN LAW INSTITUTE, UNIFORM COMMERCIAL CODE, MAY MEETING REVISIONS TO PROPOSED FINAL DRAFT No. 2 (1951).

[245] Worth Street Rules, Standard Cotton Textile Salesnote IV(2) (1941). Compare UCC § 2-609.

[246] One is reminded of those cases, often brought up when the nature of "ambiguity" for parol-evidence-rule purposes is at issue, in which within a particular trade, white means black, and so forth. See Mitchell v. Henry, 15 Ch. D. 181 (1880) ("white selvage" meaning dark selvage); Smith v. Wilson, 3 B. & Ad. 728, 110 Eng. Rep. 266 (K.B. 1832) ("thousand" meaning 1200 in the rabbit trade).

to arise? Let us assume that a case comes up in which a man has been fast-talked into signing a contract by which he will have to pay roughly three times the "value" of some goods.[247] Now let us imagine the scene when the plaintiff goes on the stand to "present evidence as to . . . [the contract's] commercial setting, purpose and effect."[248] It seems to me that the scene might go something like this:

DEFENDANT'S COUNSEL:	Mr. Greed, as President of plaintiff corporation you set that company's policy, don't you?
MAX GREED:	Certainly. I own it all.
COUNSEL:	What would you say was the purpose of the contract that you entered into with Mr. MacIver?
GREED:	To make a lot of money.
COUNSEL:	And what was the effect?
GREED:	I made a lot of money.
COUNSEL:	What would happen if you charged less?
GREED:	I'd make less money.
COUNSEL:	Mr. Greed, doesn't your conscience bother you?
GREED:	Wha?

Or instead let us consider the case of a poor, ignorant lady, with seven children, who signed a contract (pursuant to which she bought a stereo set she couldn't afford) making all goods bought from a certain seller, *whenever* bought, security for any outstanding balance owed the seller, such that on default he could take anything back, even things really already paid for.[249] That hearing might sound as follows:

BUYER'S COUNSEL:	Mr. Walker-Thomas, what is the purpose of this so-called "add-on clause"?
SELLER:	It gives us an extra hook over the buyers. Sometimes you can squeeze a little more out by repossessing some of the items bought earlier and reselling them. It makes it easier to convince the buyers that if they don't pay up they're going to get hurt bad.

[247] *Cf.* American Home Improvement, Inc. v. MacIver, 105 N.H. 435, 201 A.2d 886 (1964), discussed at length in text accompanying notes 254-66 *infra*.

[248] UCC § 2-302(2).

[249] *Cf.* Williams v. Walker-Thomas Furniture Co., 350 F.2d 445 (D.C. Cir. 1965), remanding 198 A.2d 914 (D.C. Mun. Ct. App. 1964), discussed at length in text accompanying notes 268-91 *infra*.

COUNSEL:	What is the effect of this provision?
SELLER:	It's hard to tell, but we think it helps a little.
COUNSEL:	Helps do what?
SELLER:	Helps make more money.
COUNSEL:	What about commercial setting? Does everybody use this clause?
SELLER:	How would I know? What am I, some kind of conspiracy? I guess whoever can use it uses it.
COUNSEL:	What would happen if you didn't use such a provision in your contract?
SELLER:	I'd make less money.
COUNSEL:	What if you sold only to people who could afford it?
SELLER:	I'd make *much* less money.
COUNSEL:	Doesn't your conscience bother you?
SELLER:	Wha?

Or finally, to round things off, how about some testimony in the context of the standard automobile warranty disclaimer that figured in the *Henningsen* case.[250] Since we had an oligopolistic industry involved there let us put on the stand in Greek-chorus fashion the chief executive officers of the major automobile manufacturers.

PLAINTIFF'S COUNSEL:	What is the purpose of the warranty disclaimer in your contracts?
DEFENDANTS:	It helps to increase profits.
COUNSEL:	Do you all have such a provision?
DEFENDANTS:	But of course.
COUNSEL:	Why?
DEFENDANTS:	We all like to increase profits.
COUNSEL:	What would happen if you were barred from that clause?
DEFENDANTS:	We'd increase profits some other way. We're an oligopoly, you know.
COUNSEL:	Don't your consciences bother you?
DEFENDANTS:	Beg your pardon?

[250] Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960).

The point of all this is to emphasize that the kind of technical testimony which the draftsmen apparently had in mind as being tendered pursuant to section 2-302(2) is testimony which need never be tendered by anyone defending his contract. There is no clause in a contract that is "needed" by a party; it is certainly true that no one needs an entire contract to be one-sided. It is useful for a seller to be able to refuse to ship goods any time he gets nervous about the buyer's credit.[251] It is useful for a party to protect himself against forms of increased difficulty of performance which do not rise to the level of common-law "impossibility." [252] It is even more obviously useful to disclaim warranties or to limit one's liability to essentially nothing. But not being able to provide for any of these things by contract would only take from a party one of the "edges" his lawyer had tried to get for him in his form contract. The removal of the edge would presumably find its way into the final price to be charged, if that were feasible.[253] But how does a court decide if taking that particular edge is not to be permitted if in its "commercial setting" its "purpose and effect" *is to increase a party's profits*? Is the evidence to be focused on the last few years' profit picture in the industry, or of the particular party, to see if he has been making enough money to cut down on his competitive advantages? Should there also be testimony about the particular party's competitive position vis-à-vis his competitors, to see if he can stand a few new risks? Perhaps there should, but I do find it unlikely. What seems to me more likely is that section 2-302(2), as promising as it reads, and as useful as it is for showing the conscionability of clauses that didn't look it, gives little aid to one trying to find out when a clause in a commercial contract is "unconscionable." Once again, this is because a warranty disclaimer is not "like" a remedy limitation, and both of them are not "like," say, a choice-of-law provision. Any of these clauses might well be regulated, but one cannot decide any of the questions relevant to the form of that regulation so long as one is trying not to decide a question of social policy but to flesh out an incantation.

---

251 See Worth Street Rules, Standard Cotton Textile Salesnote IV(2) (1941).

252 *Cf.* MIMEO 1941 DRAFT § 1-C, comment A(6).

253 It is amusing, in a grim sort of way, that persons in a monopolistic or oligopolistic position, who are therefore most likely to use contracts of adhesion, are simultaneously in the best position to pass on to the entire market the losses and costs which they would supposedly bear if the adhesion contracts were forbidden them.

It has also been noted that who does eventually bear the costs of shifted risks is exceedingly hard to pin down. See Calabresi, *Some Thoughts on Risk Distribution and the Law of Torts*, 70 YALE L.J. 499 (1961); Morris, *Enterprise Liabilities and the Actuarial Process—The Insignificance of Foresight*, 70 YALE L.J. 554 (1961); Note, 70 YALE L.J. 453, 455-56 (1961).

THE CASES "USING" 2-302

As the history of 2-302, and the suggested guides to its operation have now been discussed, it is time to analyze those cases in which 2-302 has so far been actually involved. Strictly speaking, only one reported case relies upon section 2-302 of the Code even as an alternative ground of holding.[254] In *American Home Improvement, Inc. v. MacIver*,[255] the plaintiff was in the business of selling and installing home improvements. It agreed with the defendant to "furnish and install 14 combination windows and one door" and "flintcoat the side walls" on defendant's home, all for $1,759.00. Since the defendant was apparently unwilling or unable to pay cash, the plaintiff undertook to arrange long-term financing, and furnished defendant with an application to a finance company (apparently in some way allied or affiliated with the plaintiff). This application was shortly "accepted," and defendant was notified in writing that his payments for the improvements would be $42.81 per month for sixty months (a grand total of $2,568.60) which included "principal, interest and life and disability insurance." Plaintiff commenced work, but after it had completed only a negligible portion of the job it was asked by defendant to stop and it complied, thereafter suing defendant for damages for breach of contract.[256]

On these facts, the New Hampshire court need never have reached any unconscionability question. There was in effect in the jurisdiction a "truth-in-lending" statute[257] which applied to the transaction. The

[254] This is not strictly true, perhaps. One Referee in Bankruptcy recently held two *security* agreements "unconscionable," purportedly relying on UCC § 2-302. Matter of Elkins-Dell Mfg. Co. (E.D. Pa.); Matter of Dorset Steel Equipment Co. (E.D. Pa.), both described at 39 REF. J. 115-16 (1965). On appeal, the District Court remanded both cases for further factual development, but declined specifically to hold § 2-302 applicable to nonsales agreements. *In re* Elkins-Dell Mfg. Co., 253 F. Supp. 864, 873 n.4 (E.D. Pa. 1966).

[255] 105 N.H. 435, 201 A.2d 886 (1964). The case has been interestingly commented upon. 78 HARV. L. REV. 895 (1965). The facts are presented as part of an agreed statement, 105 N.H. at 436-37, 201 A.2d at 886-87.

[256] The plaintiff quite properly did not claim the contract price. See UCC §§ 2-708, 2-709.

[257] N.H. REV. STAT. ANN. § 399-B:2 (Supp. 1965) forbids the extension of credit unless at the time thereof the borrower is furnished

> a clear statement in writing setting forth the finance charges, expressed in dollars, rate of interest, or monthly rate of charge, or a combination thereof, to be borne by such person in connection with such extension of credit as originally scheduled.

The court concluded reasonably enough that the requirements of this statute had not been met by the plaintiffs, and that

> In the circumstances of the present case . . . the purpose of the disclosure statute will be implemented by denying recovery to the plaintiff on its contract and granting the defendants' motion to dismiss.

105 N.H. at 439, 201 A.2d at 888.

court could have relied upon that statute to strike the contract, and indeed did so as an alternative ground of decision. But the court most specifically made it a point not to rest its decision solely upon the disclosure statute. It said:

> There is another and independent reason why the recovery should be barred in the present case because the transaction was unconscionable. "The courts have often avoided the enforcement of unconscionable provisions in long printed standardized contracts, in part by the process of 'interpretation' against the parties using them, and in part by the method used by Lord Nelson at Copenhagen." 1 Corbin, Contracts, s. 128 (1963). Without using either of these methods reliance can be placed upon the Uniform Commercial Code (U.C.C. 2-302(1)) [quotation of section omitted] . . . .
>
> Inasmuch as the defendants have received little or nothing of value and under the transaction they entered into they were paying $1,609 for goods and services valued at far less, the contract should not be enforced because of its unconscionable features. This is not a new thought or a new rule in this jurisdiction. See Morrill v. Bank, 90 N.H. 358, 365, 9 A.2d 519, 525; "It has long been the law in this state that contracts may be declared void because unconscionable and oppressive . . . ." [258]

All right, then. As of the time of writing, the only case which has relied upon section 2-302 as a basis of decision has decided that "unconscionable" means "too expensive." [259] And certainly there is no immutable principle displayed in fixed stars that would make that particular meaning of unconscionable inconceivable. I have earlier suggested that in fact that was the primary meaning of unconscionability in some of the early drafts of the Code, and that it was its *only* meaning as used by courts of equity. Certainly the idea that a strikingly dis-

---

[258] *Id.* at 439, 201 A.2d at 888-89. The *Morrill* case cited by the Court in *MacIver*, Morrill v. Amoskeag Sav. Bank, 90 N.H. 358, 9 A.2d 519 (1939), cites for the proposition quoted from it (which, by the way, is dictum) five other cases, Villa v. Rodriguez, 79 U.S. (12 Wall.) 323 (1871); Russell v. Southard, 53 U.S. (12 How.) 139 (1851); Bither v. Packard, 115 Me. 306, 98 Atl. 129 (1916); Smith v. Smith, 82 N.H. 399, 135 Atl. 25 (1926); Houghton v. Page, 2 N.H. 42 (1819). The contract in the last-named case was not enforced because it violated a governing usury statute of the *locus contractus*. All of the other cited cases (including *Morrill*) were actions in equity except the *Bither* case, which, however, was a mortgagor-mortgagee case of which the court said "though the form or proceeding is in law, it is equitable in spirit and purpose." 115 Me. at 312, 98 Atl. at 932.

[259] The only commentary devoted to *MacIver* seems to agree with that reading of its holding. The headnote to the discussion at 78 HARV. L. REV. 895 (1965) reads "Uniform Commercial Code—Construction—Inadequacy of Consideration is Sufficient To Establish Unconscionableness of Contract." Professor Braucher agrees. See Braucher, *supra* note 243, at 205, for a similar reading.

proportionate exchange should be voidable has not destroyed the commerce of the many jurisdictions which utilize a *laesio enormis* doctrine in one form or another.[260] On the other hand one may certainly speculate on whether the legislatures which have flocked to embrace the Code [261] would have been willing to adopt a provision which frankly and openly declared that overcharges of some large but unspecified degree could be invalidated by courts on an *ad hoc* basis, at least as part of a *commercial* code.

Let us assume, however, that a system of jurisprudence ought to have some way to deal with transactions in which one party is giving up vastly more than he is getting, and that this purpose is at least one of those that section 2-302 is designed to serve.[262] Even given that assumption, one has still to ask whether the best way to inject that supervisory power into the law is to subsume it under a high-level abstraction like "unconscionability." After all, a *laesio enormis* type of statute is not very hard to draft, as either a flat-percentage or a "grossly-too-much" provision. The decision in the *MacIver* case exposes the weaknesses of abstraction so deliciously that it justifies esurient consideration. Let me quote the court's total discussion of *why* the contract was unconscionable.

> In examining the exhibits and agreed facts in this case we find that to settle the principal debt of $1,759 the defendants signed instruments obligating them to pay $42.81 for 60 months, making a total payment of $2,568.60, or an increase of $809.60 over the contract price. In reliance upon the total payment the defendants were to make, the plaintiffs pay a sales commission of $800. Counsel suggests that the goods and services to be furnished the defendants thus had a value of $959, for which they would pay an additional $1,609.60 computed as follows:
>
> | | | |
> |---|---|---|
> | "Value of goods and services | | $ 959.00 |
> | Commission | 800.00 | |
> | Interest and carrying charges | 809.60 | 1,609.60 |
> | Total payment | | $2,568.60" [263] |

[260] *E.g.*, CODE CIVIL art. 1134 (Fr. 58th ed. 1959) (5/12 of value); CIVIL CODE § 138 (Ger. 10th ed. Palandt 1952) ("strikingly disproportionate"), LA. CIV. CODE ANN. art. 1861(2) (West 1952) (½ of value). Moreover, according to an exceedingly interesting recent book, the Barotse of Northern Rhodesia also know fair-price and *laesio enormis* doctrines. See GLUCKMAN, THE IDEAS IN BAROTSE JURISPRUDENCE 192-93 (1965). So much for you, Tom Hobbes.

[261] Forty-nine as of December, 1966. See 3 BENDER'S UNIFORM COMMERCIAL CODE SERVICE *vi* (Duesenberg & King eds.) (1966).

[262] It is, in fact, hypothesized in another section of the Code as one of the purposes of the whole law of warranty. See UCC § 2-313, comment 4.

[263] 105 N.H. at 438-39, 201 A.2d at 888.

This is breathtaking economics. Note first of all the court's assumption that the seller's cost of distribution (his presumably-fast-talking salesman's commission) is *no* part of the "value" of the goods to the purchaser, so that it must be *totally* eliminated from any evaluation of the fairness of the exchange. On this reasoning, the salary of sales staff is not a factor fairly to be considered when deciding the fairness of a retail price. (This is not to say that a grotesquely uneconomic form of distribution enriches one who purchases from the distributor; it does, however, suggest that life is not so easy that the commission may simply be "drilled out" of the "value" in evaluating the fairness of the exchange.)

Then we come to the treatment of the "interest and carrying charges" by the court. Here it seems that it is enough for the judge that the amount certainly *looks* like a great deal. He makes absolutely no attempt to work out what the true effective yearly rate of interest is for this five-year pay out. Actually it works out to a bit over 18% per annum.[264] This may seem high, but it is not out of line with the rates permitted under statutes which regulate installment purchases and loans, not to mention rates charged where not regulated.[265] The important matter, however, is not whether this rate is "fair" or not; it is that the court in this case went on nothing but guesswork to reach its decision, examined none of the relevant considerations *and was encouraged by 2-302 to behave in just that way.*[266] Had the section been in less abstract terms, perhaps an examination of the relevant factors would have taken place. It does seem that a judge who is forced to

[264] The formula for finding the annual simple interest rate for any time period or amount when based on a monthly repayment schedule may be expressed:

$$\frac{24\ C}{L\ (N + 1)} = R$$

where C is the cost of the loan, L is the amount of the loan, N is the number of payments to be made and R is the annual simple interest rate. See Willging, *Installment Credit: A Social Prospective*, 15 CATHOLIC U.L. REV. 45, 66 (1966) for a more generalized form of this equation. In the *MacIver* case the calculation is:

$$\frac{24 \times 810}{1759 \times 61} = 18.1\%$$

and this assumes that the "life and disability insurance" included is part of the "interest" charge. (If one gives no "value" to the salesman's commission, then the calculation yields a bit over 33%.)

[265] See CURRAN, TRENDS IN CONSUMER CREDIT LEGISLATION 69 n.558, 270-77 (1965), for the regulated rates. The first tentative draft of a Uniform Consumer Credit Code calls for a maximum rate of 18%. N.Y. Times, Aug. 5, 1966, p. 39, cols. 1 & 2. The New Hampshire statute, N.H. REV. STAT. ANN. §§ 399-B:2-8 (Supp. 1965), has no maximum charge provided. In states without rate regulation it is believed that the actual rates charged may range between 12% and 275%. See Fand, *Competition and Regulation in the Consumer Credit Markets*, 20 PERSONAL FINANCE L.Q. REP. 18, 23 (1965).

[266] In fairness, it should be pointed out that the plaintiff-appellee filed no brief at all on appeal to the New Hampshire Supreme Court. See 105 N.H. at 437, 201 A.2d at 887.

recognize by the form of the statute upon which he is relying that he is supposed to evaluate the actual economic exchange is likely to feel called upon to see what that exchange in fact is, and how it accords with contemporary practices. When the key evaluative word, however, is a description of the judge's own state of mind rather than of the situation which might be justified in producing such a state, the likelihood that the court will even examine the relevant questions is severely lessened.[267]

As noted earlier, the *MacIver* case is the only one reported which has relied upon 2-302 as a basis of decision. One very recent case, however, which has attracted substantial attention from the commentators, clearly would have been decided on the basis of 2-302 had the statute been in effect at the time of the relevant transaction, and in fact was decided as if the section were the law of the jurisdiction. In that case, *Williams v. Walker-Thomas Furniture Co.*,[268] the appellant, a Mrs. Williams, "a person of limited education separated from her husband," [269] had, during the period 1957-1962, purchased "a number of household items" [270] from appellee furniture company on printed-form installment sale contracts (in the transparent guise of leases). One sentence in this printed contract, part of "a long paragraph in extremely fine print" [271] had the net effect of keeping

> a balance due on every item purchased until the balance due on all items, whenever purchased, was liquidated. As a result, the debt incurred at the time of purchase of each item was secured by the right to repossess all the items previously purchased by the same purchaser . . . .[272]

When Mrs. Williams' outstanding balance was only $164, she bought a $515 stereo phonograph set.[273] At the time of this purchase, the

[267] The court's *sole* reference, by the way, to the procedural unconscionability problems is the following quote from Corbin:

> "The courts have often avoided the enforcement of unconscionable provisions in long printed standardized contracts, in part by the process of 'interpretation' against the parties using them, and in part by the method used by Lord Nelson at Copenhagen." 105 N.H. at 439, 201 A.2d at 888.

[268] 350 F.2d 445 (D.C. Cir. 1965), *remanding* 198 A.2d 914 (D.C. Mun. Ct. App. 1964). The case has occasioned extensive commentary. See, *e.g.*, Schneider, *Unconscionable Contract Unenforceable*, 20 PERSONAL FINANCE L.Q. REP. 32 (1965); 15 CATHOLIC U.L. REV. 269 (1966); 54 GEO. L.J. 703 (1966); 79 HARV. L. REV. 1299 (1966); 12 How. L.J. 164 (1966); 44 TEXAS L. REV. 803 (1966).

[269] 198 A.2d at 915.

[270] This abstract classification is that of the Court of Appeals. 350 F.2d at 447. The lower court opinion is, interestingly enough, not as reticent, spelling out the items as "sheets, curtains, rugs, chairs, a chest of drawers, beds, mattresses, a washing machine." 198 A.2d at 915.

The *Williams* case had a companion case involving a Mr. Thorne and Walker-Thomas, which was also remanded. Mr. Thorne had purchased instead of a stereo set, "an item described as a Daveno, three tables, and two lamps." 350 F.2d at 447. Mr. Thorne's marital status and number of children are not described.

[271] 198 A.2d at 915.

[272] 350 F.2d at 447.

[273] 350 F.2d at 447 & n.1.

furniture company was perfectly aware (since the information was endorsed on the back of the installment contract) that Mrs. Williams' sole income [274] was a government payment (apparently some species of relief) of $218 per month. The Circuit Court opinion also indicated that the store knew that Mrs. Williams was supposed to support herself and her seven children on that amount (though that seems not to have been endorsed on the back of the contract).[275] At any rate, the stereo set was apparently just too great a burden for the $218 per month to bear. Mrs. Williams defaulted, the store replevied every item it could lay its hands on and won in the trial court and the intermediate appeals court. On appeal to the United States Court of Appeals for the District of Columbia Circuit, the case was remanded to make findings on the issue of unconscionability.[276]

For those of us who have an instinctive and infallible sense of justice (and which of us does not), any other result in this case is unimaginable. But there are grounds for quibbling about the court's (and the Code's) methodology. Judge Wright found unconscionability easy to describe:

> Unconscionability has generally been recognized to include an absence of meaningful choice on the part of one of the parties together with contract terms which are unreasonably favorable to the other party. . . . [In the footnote which supports this statement, citation is to Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960), and Campbell Soup Co. v. Wentz, 172 F.2d 80 (3d Cir. 1948) only.] In many cases the meaningfulness of the choice is negated by a gross inequality of bargaining power.[277]

That is, there is immediate recognition that unconscionability has to have two foci, the negotiation which led to the contract and that contract's terms. As for the procedural aspect, while there is no finding that this was the only credit furniture store open to Mrs. Williams, or that even if it were not, they all had substantially the same contract (which was the situation in the *Henningsen* case [278] so heavily relied

---

[274] Actually, there is no finding that Mrs. Williams had no other source of income, but one assumes that if she had, one of the courts, most likely the Municipal Court of Appeals, would have mentioned it.

[275] See 350 F.2d at 448. Presumably after fourteen contracts over a five-year period, see 12 How. L.J. 164, 168-69 (1966), Mrs. Williams' home status was the subject of common banter among the Walker-Thomas folks.

[276] 350 F.2d at 450.

[277] *Id.* at 449. Compare § 2-302, comment 1: "The principle [of the section] is . . . not [one] of disturbance of allocation of risks because of superior bargaining power."

[278] Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960).

upon by the court), one may assume that the form Mrs. Williams signed was essentially the only kind of form open to her. A person's "relevant market" may fairly be the one he can reasonably be expected to know about or dare to use. In other words, the local stores may be a local person's relevant market because of his ignorance, and if they are all as one on something, as to him they are a monopoly. And besides, in this case the court made an almost-finding of contracting procedures which went beyond the mere use of a form or even of a contract of adhesion, which reached, in fact, at least some level of quasi-fraud. Judge Wright asks:

> Did each party to the contract, considering his obvious education or lack of it, have a reasonable opportunity to understand the terms of the contract, or were the important terms hidden in a maze of fine print and minimized by deceptive sales practices? [279]

There was apparently no problem with the answer, for after giving lip service to the "usual rule" [280] that one who signs an agreement is bound to all of its terms, he said:

> But when a party of little bargaining power, and hence little real choice, signs a commercially unreasonable contract with little or no knowledge of its terms . . . . the usual rule . . . should be abandoned . . . .[281]

It is hard to fault the court's argument on the procedural unconscionability aspects of this case. While it might sometimes be difficult to decide whether a species or level of bargaining ought to protect a contract from section 2-302, it is not difficult here. If the unconscionability section is to be applicable to any contract, it must be to one "bargained" as this one was.

But there is no need to labor this point. Finding that the bargaining procedure involved will not insulate the contract from judicial scrutiny under section 2-302 is only the first and less difficult step in the process of using that section. Once one decides that the contract is vulnerable to judicial meddling, there still remains to be decided whether the provision or contract is "unconscionable." For that determination Judge Wright also articulated a test:

> In determining reasonableness or fairness, the primary concern must be with the terms of the contract considered

---

279 350 F.2d at 449.

280 See RESTATEMENT, CONTRACTS § 70 (1932).

281 350 F.2d at 449. Compare § 2-302, comment 1: "The principle [of the section] is . . . not [one] of disturbance of allocation of risks because of superior bargaining power."

> in light of the circumstances existing when the contract was made. The test is not simple, nor can it be mechanically applied. The terms are to be considered "in the light of the general commercial background and the commercial needs of the particular trade or case" [citing "Comment, Uniform Commercial Code sec. 2-307," but obviously meaning 2-302]. Corbin suggests the test as being whether the terms are "so extreme as to appear unconscionable according to the mores and business practices of the time and place." . . . We think this formulation correctly states the test to be applied in those cases where no meaningful choice was exercised upon entering the contract.[282]

How does that test apply to the *Williams* facts? What is it about Mrs. Williams' contract which is "unconscionable"? Surprisingly, the answer is not clear, even about *what* in the contract is bad. It seems, however, that there are two possibilities. First, it may be that the provision by which each item purchased became security for all items purchased was the objectionable feature of the contract. Or it might be that the furniture company sold this expensive stereo set to this particular party which forms the unconscionability of the contract. If the vice is the add-on clause, then one encounters the now-familiar problem: such a clause is hardly such a moral outrage as by itself meets Judge Wright's standard of being "so extreme as to appear unconscionable according to the mores and business practices of the time and place." The lower court in the *Williams* case called attention,[283] for instance, to a Maryland statute regulating retail installment sales [284] under which Mrs. Williams might have been relieved, noting with regret that the statute was not in effect in the District. What was not pointed out by the lower court (and *certainly* not by the upper court) was that the State of Maryland had found nothing illegal per se about add-on provisions, in fact specifically permitting them and setting out to regulate them in some detail.[285] Of the thirty-seven jurisdictions which have statutes regulating retail installment sales, only one has a provision making add-on clauses impermissible.[286] In such circumstances it does seem a bit much to find "so extreme as to appear unconscionable according to the mores and business practices of the time

[282] 350 F.2d at 450.

[283] 198 A.2d at 916.

[284] MD. ANN. CODE art. 83, §§ 128-53 (1965).

[285] MD. ANN. CODE art. 83, § 137 (1965).

[286] As of 1965. See CURRAN, TRENDS IN CONSUMER CREDIT LEGISLATION 312-22 (1965). A bill proposed for the District of Columbia would apparently at least regulate the use of the add-on provisions. See 12 How. L.J. 164, 172 (1966).

and place" [287] an add-on clause in the District of Columbia which is used and statutorily permitted almost everyplace else, including contiguous Maryland. One's gorge can hardly be expected to rise with such nice geographic selectivity.

If one is not convinced that the unconscionability inheres in the add-on provision, it may be argued that it inheres in the contract as a whole, in the act of having sold this expensive item to a poor person knowing of her poverty. This is quite clearly the primary significance of the case to some of the commentators.[288] That is the kind of action which the Maryland statute does not deal with, nor do any of the statutes like it: the unconscionability of aiding or encouraging a person to live beyond his means (without much hope of eventual success). Well, why not make that "unconscionable" for purposes of section 2-302? After all, in this case Walker-Thomas did know for a fact that Mrs. Williams was on relief; they knew her income and needs with great particularity: $218 per month and seven children. This case does not present any of the sticky close questions of how much of what a seller would have to know (or inquire about) before being deemed to know that the buyer shouldn't buy.[289] Moreover, what Mrs. Williams bought this time was a stereo record player. No one could argue that such an article is a "necessity" to a relief client, and thus the dissenting judge's suggestion that "what is a luxury to some may seem an outright necessity to others" [290] hardly applies in this case. Who can doubt but that this purchase was a frill? [291] So in this case all we would have is a holding that one cannot enforce a contract pursuant to which one has sold luxuries to a poor person (or at least one on relief) with knowledge or reason to know that he will not be able to pay for them. This is just another class distinction, and distinctions among persons on the bases of the "class" to which they belong, that is, on the basis of some common supra-personal charac-

---

287 350 F.2d at 450.

288 For instance, the subheading of Schneider, *Unconscionable Contract Unenforceable*, 20 PERSONAL FINANCE L.Q. REP. 32 (1965), is "Sale of Stereo to Woman on Relief." See also 44 TEXAS L. REV. 803 (1966): "a court may refuse on the ground of unconscionability to enforce a contract where there is an overextension of credit." The great interest shown in the opinion as to Mrs. Williams' financial status and the prominence given her relief status seems to express a similar feeling about the case's real significance.

289 For simplicity's sake we will leave Mr. Thorne out of this. See note 270 *supra*.

290 350 F.2d at 450.

291 *But see* LEWIS, FIVE FAMILIES 134-36 (1959), for the story of a Mexican slum dweller who bought a "combination radio, record player, and television set" on time, intending to rent its use to his neighbors until it was paid for and then (having left it carefully cartoned and protected) resell it as new. One is not told how the plan finally worked out.

teristics, is exceedingly common in the law [292] (not to mention life). Such a process immensely simplifies decision by limiting the required inquiry to the person's membership in the class. Once that determination is made, a certain legal result will flow. The classic instance is the majority-minority dichotomy. Persons under twenty-one cannot, as a general rule, make self-binding contracts. This may be considered a shorthand form of a syllogism which would go something like (a) persons lacking sufficient probity ought not to be allowed to bind themselves; (b) all persons under twenty-one lack sufficient probity; (c) persons under twenty-one cannot bind themselves. This illustrates some of the strengths and weakneses of the class system. The rule is easy to administer because a party's age is much easier to determine than his probity. The difficulty is that the easier the classification the less likely it is to be accurate, because classes are, in fact, hardly ever wholly homogeneous. In our case, for instance, the "minor premise" is false; not all persons under twenty-one lack sufficient probity to bind themselves.[293]

When faced with the difficulties inherent in deciding the bargaining fairness of any given transaction, the equity courts, in working out *their* unconscionability doctrine, similarly leaned heavily on relatively gross classifications. In effect, they seem continually to have taken a kind of *sub rosa* judicial notice of the amount of power of certain classes of people to take care of themselves, often without too much inquiry into the actual individual bargaining situation. And it is arguable that sometimes they were wrong; not all old ladies or farmers are without defenses.[294] Put briefly, the typical has a tendency to become stereotypical, with what may be unpleasant results even for the beneficiaries of the judicial benevolence. One can see it enshrined in the old English equity courts' jolly treatment of English seamen as members of a happy, fun-loving race (with, one supposes, a fine sense of rhythm), but certainly not to be trusted to take care of themselves.[295] What effect, if any, this had upon the sailors is hidden behind the judicial chuckles

---

[292] The usefulness of these group classifications has not gone unnoticed by the draftsmen of the Code, whose innovations, after all, include the merchant-non-merchant dichotomy, § 2-104(1). See also Llewellyn, *Through Title To Contract and a Bit Beyond*, 15 N.Y.U.L. REV. 159, 160 n.2 (1938): "My own attack would be to . . . split 'retailer' still further into petty, and large (department store; chain store)."

[293] For some recent speculations on infants' contracts in the modern world, see Note, 41 IND. L.J. 140 (1965).

[294] For instance, consider farmer Wentz in Campbell Soup Co. v. Wentz, 172 F.2d 80 (3d Cir. 1948).

[295] See How v. Weldon, 2 Ves. Sr. 516, 518, 28 Eng. Rep. 330, 331 (Ch. 1754) ("a race of men, loose and unthinking"); 3 POMEROY, EQUITY JURISPRUDENCE § 952 (5th ed. Symons 1941); 1 STOREY, EQUITY JURISPRUDENCE § 332 (13th ed. Bigelow 1886).

as they protected their loyal sailor boys, but one cannot help wondering how many sailors managed to get credit at any reasonable price. In other words, the benevolent have a tendency to colonize, whether geographically or legally.[296]

Far more economically significant and widespread as an example of the Chancellors' temptation toward stereotypical jurisprudence is found in the expectant-heir cases. The most important thing about expectant-heir cases is that there are expectant-heir cases, classifiable separately as such in treatises.[297] The Chancellors did not find unfairness in the price and refuse to enforce because they had no conception of how an expectancy had to be discounted for risk; that kind of sophistication came early.[298] They just set out to protect heirs from the full effect of their tendency to live beyond their governors' life expectancies. This was easy to do; it was rare that a judge had to enter into too long a discussion of the actual facts, or to face the real basis of his easy decision in the battle between his (there but for the grace of God) grandson and the most-likely-Jewish moneylender. After all, he had the rubric "unconscionable" with which to explain (to himself and to the public) that decision.

Thus, when one asks why a court (like the District of Columbia Court in the *Williams* case) ought not be allowed to subsume its social decisions under a high-level abstraction like "unconscionability," one may point to the equity cases so many other commentators have pointed to, but for a different reason. One may suggest that first (and less important) it tends to make the true bases of decisions more hidden to those trying to use them as the basis of future planning. But more important, it tends to permit a court to be nondisclosive about the basis of its decision even to itself; the class determination is so easy and so

---

296 See F. Cohen, *Indian Wardship: The Twilight of a Myth*, in THE LEGAL CONSCIENCE 328 (L. Cohen ed. 1960); *cf.* Williams v. Walker-Thomas Furniture Co., 350 F.2d 445 (D.C. Cir. 1965).

297 See FONBLANQUE, A TREATISE OF EQUITY 119-22 (4th Am. ed. 1835); 3 POMEROY, EQUITY JURISPRUDENCE §§ 953-54 (5th ed. Symons 1941); POMEROY, SPECIFIC PERFORMANCE OF CONTRACTS § 191 (3d ed. 1926); 1 STORY, EQUITY JURISPRUDENCE §§ 333-34 (13th ed. Bigelow 1886). In 1804 Lord Eldon could distinguish the "cases of reversions and interests of that sort" as a separate class of cases, Coles v. Trecothick, 9 Ves. Jr. 234, 246, 32 Eng. Rep. 592, 597 (Ch. 1804), and modern English treatises continue to note them separately. See HORSFIELD, EQUITY IN A NUTSHELL 181 (1960); SNELL, PRINCIPLES OF EQUITY 502-03 (25th ed. Megarry & Baker 1960). In fact, cases are still coming up, *e.g.*, Levin v. Roth, [1950] 1 All E.R. 698 (C.A.), and the old expectant-heir learning is still discussed. 24 AUSTL. L.J. 158 (1950):

> Though of somewhat rare occurrence before the Courts, the application of the equitable doctrine relating to unconscionable bargains with expectant heirs should not be forgotten.

298 The leading case of Earl of Chesterfield v. Janssen, 2 Ves. Sr. 125, 28 Eng. Rep. 82 (Ch. 1750), for instance, contains a sophisticated judicial discussion of the economic problems involved.

tempting (and often so heart-warming). More particularly with respect to the *Williams* case concept that the poor should be discouraged from frill-buying, no legislature in America could be persuaded openly to pass such a statute, nor should any be permitted to do so sneakily. If the selling of frills to the poor is to be discouraged, if the traditional middle-class virtues of thrift and child care are to be fostered in the deserving poor by a commercial statute, if one wants to protect a class, improvident by definition, from the depredations of another class, it is at least arguable that one should just up and do so—but clearly.[299] This is not to suggest, for a moment, anything as stupid as that some "freedom-of-contract" concept ought to prevent, for instance, the statutory interdiction of an eleven-hour day. It is only to say that when you forbid a contractual practice, you ought to have the political nerve to do so with some understanding (and some disclosure) of what you are doing.[300]

CONCLUSION

I have attempted to describe in some detail the pathology, developmental, morphological and functional, of section 2-302 of the Code and its official and unofficial commentaries. The gist of the tale is simple: it is hard to give up an emotionally satisfying incantation, and the way to keep the glow without the trouble of the meaning is continually to increase the abstraction level of the drafting and explaining language. If for one reason or another (in this case the desire to forward the passage of the whole Code) the academic community is generally friendly to the drafting effort, a single provision in a massive Code may get by even if it has, really, no reality referent, and all of its explanatory material ranges between the irrelevant and the misleading. That this happened with respect to 2-302 the few cases using it are beginning to show more and more clearly. The world is not going to come to an end. The courts will most likely adjust, encrusting the irritating aspects of the section with a smoothing nacre of more or less reasonable applications, or the legislatures may act if things get out of

[299] See 12 How. L.J. 164, 170-72 (1966), for a brief contemplation of such overt class legislation which is notably lacking in enthusiasm. At least one observer of the consumer finance scene, however, seems willing to bar the poor from credit. See CAPLOVITZ, THE POOR PAY MORE 191 (1963) ("establish by law minimal credit requirements").

[300] In Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960), continually cited in the *Williams* case (see 350 F.2d at 448 n.2, 449 nn.6 & 7, 450 n.12), it is most significant that the court did *not* have § 2-302 to work with. It was forced, therefore, to face the relevant policy questions, which it did in a many-paged opinion. In other words, in *Henningsen* the New Jersey court was forced to talk about the basis for its decision; in *Williams* and *MacIver* the courts were most particularly enabled not to.

hand. Commerce in any event is not going to grind to a halt because of the weaknesses in 2-302. But the lesson of its drafting ought nevertheless to be learned: it is easy to say nothing with words. Even if those words make one feel all warm inside, the result of sedulously preventing thought about them is likely to lead to more trouble than the draftsmen's cozy glow is worth, as a matter not only of statutory elegance but of effect in the world being regulated. Subsuming problems is not as good as solving them, and may in fact retard solutions instead. Or, once more to quote Karl Llewellyn (to whom, after all, the last word justly belongs), "Covert tools are never reliable tools." [301]

[301] LLEWELLYN, THE COMMON LAW TRADITION 365 (1960), quoting Llewellyn, Book Review, 52 HARV. L. REV. 700, 703 (1939).

EMPLOYMENT AT WILL VS. INDIVIDUAL FREEDOM: ON LIMITING THE ABUSIVE EXERCISE OF EMPLOYER POWER

LAWRENCE E. BLADES*

It is a widely accepted proposition that large corporations now pose a threat to individual freedom comparable to that which would be posed if governmental power were unchecked.[1] The proposition need not, however, be limited to the mammoth business corporation, for the freedom of the individual is threatened whenever he becomes dependent upon a private entity possessing greater power than himself.[2] Foremost among the relationships of which this generality is true is that of employer and employee.[3]

The threat to individual freedom posed by employer power has special significance because

> We have become a nation of employees. We are dependent upon others for our means of livelihood, and most of our people have become completely dependent upon wages. If they lose their jobs they lose every resource, except for the relief supplied by the various forms of social security. Such dependence of the mass of the people upon others for *all* of their income is something new in the world. *For our generation, the substance of life is in another man's hands.*[4]

It is well known that the labor union movement was a response to the

---

* Associate Professor of Law, The University of Kansas. A.B., Dartmouth, 1957; J.D., Michigan, 1960.

1. [N]ot only do 500 corporations control two-thirds of the non-farm economy but within each of that 500 a still smaller group has the ultimate decision-making power. This is, I think, the highest concentration of economic power in recorded history.

A. BERLE, ECONOMIC POWER AND THE FREE SOCIETY 14 (Fund for the Republic, 1957). The gross revenues of each of [General Motors, Standard Oil of New Jersey and Ford Motor Company] far exceed those of any single state. The revenues of General Motors in 1963 were fifty times those of Nevada, eight times those of New York and slightly less than one-fifth those of the Federal Government.

J. K. GALBRAITH, THE NEW INDUSTRIAL STATE 76 (1967).

Kaysen, *The Corporation: How Much Power? What Scope?*, in THE CORPORATION IN MODERN SOCIETY 85 (E. Mason ed. 1960) describes the "proposition that a group of giant business corporations, few in number but awesome in aggregate size, embodies a significant and troublesome concentration of power" as a "cliche," but is careful to point out that the familiarity of the proposition "is no argument against its truth." *See, e.g.*, Berle, *Constitutional Limitations on Corporate Activity—Protection of Personal Rights from Invasion Through Economic Power*, 100 U. PA. L. REV. 933 (1952); Latham, *The Commonwealth of the Corporation*, 55 NW. U. L. REV. 25 (1960).

2. It has been pointed out that the problem of group power involves not only large corporations, but labor unions, charitable foundations, and other organized private groups as well. Friedmann, *Corporate Power, Government by Private Groups, and the Law*, 57 COLUM. L. REV. 155 (1957).

3. Government employment, and the considerations peculiar to the public employee, are beyond the scope of this discussion. *See generally* Emerson & Helfeld, *Loyalty Among Government Employees*, 58 YALE L.J. 1 (1948); Dotson, *The Emerging Doctrine of Privilege in Public Employment* 15 PUB. AD. REV. 77 (1955); Note, *Dismissal of Federal Employees—The Emerging Judicial Role*, 66 COLUM. L. REV. 719 (1966); Ridgeway, *The Snoops, Private Lives and Public Service*, THE NEW REPUBLIC, Dec. 19, 1964, p. 13.

4. F. TANNENBAUM, A PHILOSOPHY OF LABOR 9 (1951) (emphasis in original).

imbalance in the relationship of the individual employee to his employer. Labor unionism, however, has only partially filled the need to lessen employer-employee inequality. In the many employment relationships not covered by collective bargaining agreements, much the same imbalance which produced unionism still exists. And despite the aggravation of this imbalance by the ever-increasing concentration of economic power in the hands of fewer employers,[6] the law has done little, outside the limited and shrinking realm of labor unions,[7] to protect the economically dependent employee from employer power.

I. The Absolute Right of Discharge—The Prime Source of the Employer's Power over His Employee

Obviously, if every employee could go from job to job with complete ease, there would be little need to provide other means of escape from the improper exertion of employer pressure. In reality, however, "[b]ecause of his comparative immobility, the individual worker has long been highly vulnerable to private economic power."[8] This immobility is being enhanced by the narrowing of the range of alternative employment as advances in modern technology require more and more specialization.[9] Thus the prospect is that concern over job security will increase, and that employees therefore will become even more easily oppressed by their employers.

Despite this irreversible trend, the law has adhered to the age-old rule that all employers "may dismiss their employees at will . . . for good cause, for no cause, *or even for cause morally wrong,* without being thereby guilty of legal wrong."[10] This traditional rule, which forces the non-union employee to rely on the whim of his employer for preservation of his livelihood, is what most tends to make him a docile follower of his employer's every wish. There are, to be sure, less drastic threats than that of discharge by which an employer might bend the will of his employee to his own. An employee who

5. *See* American Steel Foundries v. Tri-City Central Trades Council, 257 U.S. 184, 209 (1921) (Taft, C.J.):

> A single employee was helpless in dealing with an employer. He was dependent ordinarily on his daily wage for the maintenance of himself and family. If the employer refused to pay him the wages that he thought fair, he was nevertheless unable to leave the employ and to resist arbitrary and unfair treatment. Union was essential to give laborers opportunity to deal on equality with their employer.

6. The certainty that the United States will have one hundred million more people thirty years from now guarantees that there will be more, not less, organized economic power in the future.
Berle, *Legal Problems of Economic Power,* 60 Colum. L. Rev. 4, 11 (1960).

7. See text accompanying notes 30-37 *infra.*

8. J. K. Galbraith, American Capitalism 114 (2d ed. 1956).

9. For an extreme example see B. F. Goodrich v. Wohlgemuth, 137 U.S.P.Q. 804 (Ohio App. 1963). Having learned valuable trade secrets while employed as a space suit expert by the plaintiff company, the employee was enjoined from going to work for one of the few other companies in the space suit field.

10. Payne v. Western & A. R.R., 81 Tenn. 507, 519-20 (1884), *overruled on other grounds,* Hutton v. Watters, 132 Tenn. 527, 179 S.W. 134 (1915) (emphasis added). *See also* 1 C. Labatt, Master and Servant § 183 (2d ed. 1904).

balks at an employer's demand might, for instance, have his compensation reduced, be given unfavorable work assignments, be subjected to the inconvenience of frequent transfers, or be foreclosed from any hope of advancement. Such harassment can be aggravated to a point where it is the practical equivalent of discharge and should be treated as such. But ordinarily the employee might be expected to bear the pressure of unfavored treatment, so long as he is able to maintain a steady income while perhaps searching discreetly for other suitable employment. It is quite another thing, however, to expect the employee to risk having his present job pulled out from under him, and having the blemish of dismissal reduce his chances of finding another one. It is the fear of being discharged which above all else renders the great majority of employees vulnerable to employer coercion.[11]

II. Defining the Scope of the Employer's Appropriate Control

Certainly, the employee can never expect to be completely free to do as he pleases. He must face the prospect of discharge for failing or refusing to do his work in accordance with his employer's directions; such control by the employer over the employee is fundamental to the employment relationship.[12] On the other hand, there are innumerable facets of the employee's life that have little or no relevance to the employment relationship, and over which the employer should not be allowed to exercise control. No certain line of demarcation can be drawn, however, between the reasonable demands of an employer and those which are overreaching, for some argument can almost always be made that the employer has an interest in whatever his employee does or believes. A salesman, for example, might be told to join certain social organizations and develop acquaintances upon which business relationships might be built. Such a requirement is not uncommonly accepted as affecting a matter of legitimate concern to the employer and thus "part of the job." Then, too, depending on the employer's business and employee's position, an employer may well be legitimately concerned with many aspects of the "off hours" behavior of the employee. There may even be occasions when an employee's public utterances on controversial subjects can be considered incompatible with his professional position and the duty of loyalty he owes to his employer.[13]

---

11. A dismissal has been called "a kind of organizational equivalent of capital punishment." W. E. Moore, The Conduct of the Corporation 28 (1962).

12. [T]he relation of master and servant . . . cast certain duties upon . . . the servant, which he was bound to fulfill and discharge; and the principle [sic] one was that of obedience to all reasonable orders of . . . the master, not inconsistent with the contract. Disobedience of reasonable orders is a violation of law which justifies . . . the peremptory discharge of the servant.

Mair v. Southern Minn. Broadcasting Co., 226 Minn. 137, 138, 32 N.W.2d 177, 178 (1948), *quoting from* Von Heyne v. Tompkins, 89 Minn. 77, 81, 93 N.W. 901, 903 (1903).

13. Compare the distinction drawn in the ACLU statement entitled Academic Freedom and Academic Responsibility between "(1) the conduct of a teacher apart from

Nevertheless, the impossibility of defining with precision the scope of the employer's appropriate control over the employee is insufficient reason for treating that control as boundless. In many instances the business interests of the employer and the personal rights of the employee will be delicately balanced. The difficulty in drawing a line might warrant conceding much that is arguable. But numerous demands an employer might make of his employee, when weighed against the interests of the employee as an individual, are clearly not justified by the employer's legitimate concerns. The employee in a free society ought to be protected at least from such unquestionably overreaching domination.

To catalog in advance all the various facets of an employee's life which ought to be immune from intrusion by the employer would be impossible. Such a list could only be supplied, as suggested hereinafter, through a process of continuing judicial elaboration. But a few illustrations will give some idea of the demands which should be treated as clearly beyond the employer's legitimate concern. As suggested by the analogy which some have drawn between governmental and private power, many of the rights and privileges which are considered so important to a free society that they are constitutionally protected from government encroachment are vulnerable to abuse through an employer's power. An employer might use a threat of discharge, for example, to impair an employee's freedom against self-incrimination,[14] his political free choice[15] or his right to speak out on the issues of the day.[16]

specifically professional responsibilities and (2) his conduct in teaching and other activities directly related to professional responsibilities." 42 AM. ASS'N OF U. PROF. BULL. 517, 523 (1956). It seems that a private employee's responsibility to his employer and his freedom to speak out on the issues of the day could be reconciled along the lines suggested in the following statement on the teacher's freedom to express himself as a citizen and his responsibility to his employing institution:

> [The teacher] is not required because of his profession to maintain a timorous silence as a price of professional status . . . . However, since the public may judge his profession and his institution by his utterances, he should make every effort to maintain high professional excellence and at the same time to indicate that he does not speak for the institution which employs him. When he speaks or writes as an individual he should be free from . . . institutional . . . censorship or discipline.

42 AM. ASS'N OF U. PROF. BULL. at 518.

14. *See, e.g.*, Electrical Radio Workers v. General Electric Co., 127 F. Supp. 934 (D.D.C. 1954), *cert. denied*, 352 U.S. 872 (1956). *Cf.* Garrity v. New Jersey, 385 U.S. 493 (1967), where police officers, upon being given a choice between incriminating themselves and forfeiting their jobs, chose to confess their participation in the fixing of traffic tickets. The question presented to the Court was "whether . . . the fear of being discharged . . . for refusal to answer on the one hand and the fear of self-incrimination on the other was 'a choice between the rock and the whirlpool' which made the statements products of coercion in the violation of the Fourteenth Amendment." *Id.* at 496. The confessions so obtained were held to be coerced and thus inadmissible in subsequent criminal proceedings against the police officers, even though the confessions were obviously relevant to the proper performance of their duties. Since it is state action which the fourteenth amendment protects, the employees would not have been entitled to a similar remedy had the employer been private rather than governmental.

15. *See, e.g.*, Mims v. Metropolitan Life Ins. Co., 200 F.2d 800 (5th Cir. 1952), *cert. denied*, 345 U.S. 940 (1953); Bell v. Faulkner, 75 S.W.2d 612 (Mo. App. 1934). *Cf.* Fort v. Civil Service Comm'n, 38 Cal. Rptr. 625, 392 P.2d 385 (1964), where the plaintiff became chairman of a committee to re-elect the governor and was fired pursuant to a county charter provision prohibiting a county employee from taking part in political management

A threat to an employee's job might also secure his unwilling participation in almost any kind of immoral[17] or unlawful[18] activity. Proof of this may be found in the oft-heard apology among high-salaried executives that "my management has required, in effect, that I go along with certain antitrust violations involving restraint of trade."[19] Though the law is not indulgent of the employee who engages in criminal activity at the behest of his employer,[20] it must be recognized a substantial element of economic duress is often present —that the executive, while responsible in the eyes of the law, may really have very little freedom of choice in implementing his employer's decision to fix prices.

Another danger posed by the employer's superior power is summed up in the observation that "[t]he professional employee encounters many situations where his code of ethics comes in conflict with the requirements of his employer, where the tenets and ethics of his profession are pitted directly against his daily bread."[21] Consider, for example, the plight of an engineer who is told that he will lose his job unless he falsifies his data or conclusions, or unless he approves a product which does not conform to specifications or meet minimum standards.[22] Consider also the dilemma of a corporate attorney

---

or affairs in any political campaign. The provision was held unconstitutional in that it unreasonably abridged "fundamental rights' of county employees. Here again, as in Garrity v. New Jersey, 385 U.S. 496 (1967), the employee would not have been protected if the employer had been private rather than governmental.

Where the employee has an enforceable contract of employment for a definite time, it is generally held that he cannot be discharged for his political views unless his political activities interfere with the performance of his duties or unless his political conduct or affiliation favors a party generally held in disfavor (*e.g.*, the Communist Party) and is found thus to reflect unfavorably on the employer's interests. *See generally* Annot., 51 A.L.R.2d 742, 747 (1957).

16. Compare the system of academic tenure which has been constructed in order to protect the teacher in higher education from interference with his freedom in formulating and expressing his views. *See generally*, C. BYSE & L. JOUGHIN, TENURE IN AMERICAN HIGHER EDUCATION: PLANS, PRACTICES, AND THE LAW (1959); SYMPOSIUM, ACADEMIC FREEDOM, 28 LAW & CONTEMP. PROB. 429 (1963). The importance of academic freedom has been defended on the theory that it produces members of society who will be willing "to play the role of innovator . . . so that as adults they can more readily lay aside cultural patterns and produce ones appropriate to the times." O. BRIM, SOCIOLOGY AND THE FIELD OF EDUCATION 16 (1958). It seems that this objective of academic freedom is frustrated to the extent that one-time students, upon becoming employees, fall under the control of their employers and lose their freedom of individual expression. Particularly intolerable in a society so dependent on free, open and intelligent discussion is the silencing through their employer's power over them, of professional and other high ranking, usually non-unionized, employees.

17. *Cf.* Comerford v. International Harvester Co., 235 Ala. 376, 178 So. 894 (1938) (employee alleged that he was fired as a result of his superior's failure to "alienate the affections" of his wife).

18. *See, e.g.*, Susnjar v. United States, 27 F.2d 223 (6th Cir. 1928) (employee required by employer to smuggle aliens into country in violation of Immigration Act); Hardy v. United States, 256 F. 284 (5th Cir.), *cert. denied*, 250 U.S. 659 (1919) (employee required by employer to illegally transport liquor).

19. *See* Baumhart, *How Ethical Are Businessmen?*, HARV. BUS. REV., July-August, 1961, at 6, 164.

20. "One's participation in a criminal conspiracy is not excused by showing that the service he was employed to render required or called for such participation." Hardy v. United States, 256 F. 284, 288 (5th Cir.), *cert. denied*, 250 U.S. 659 (1919).

21. Jolly, *Needed: Stronger Engineering Groups*, CHEMICAL ENGINEERING, Aug. 3, 1964, at 112, 114.

22. There appears to be ample evidence that employee-engineers are often confronted

who is told, say in the context of an impending tax audit or antitrust investigation, to draft backdated corporate records concerning events which never took place or to falsify other documents so that adverse legal consequences may be avoided by the corporation;[23] and the predicament of an accountant who is told to falsify his employer's profit and loss statement in order to enable the employer to obtain credit.[24]

The employee also might be forced by threat of discharge not to give testimony unfavorable to the employer,[25] or to give up a lawful claim against the employer,[26] a fellow employee,[27] or some third party,[28] or not to buy goods from or otherwise deal with a particular business concern.[29]

---

with such conflicts between their professional standards and the demands of management. For example:

> I recall the dilemma of an older P.E., in the shadow of a comfortable retirement, who was confronted by a new general manager of the plant in which he was employed as a facilities engineer. In consideration of plans for a plant expansion, the General Manager insisted that the P.E. reduce footings and structural steel specifications below standards of good practice. The P.E. was told to choose between his job and his seal on the plans.

Howard, *A Bill of Professional Rights for Employed Engineers?*, AMERICAN ENGINEER, Oct., 1960, at 47, 48.

> One situation that occurs repeatedly is the question of twisting data and conclusions: Engineer Smith is directed by his supervisor to make a study of a project, with dollar estimates of work to be performed. Smith does the job as best as he can and presents the rough draft for approval prior to final preparation. The supervisor . . . directs Smith to change his estimates in a fashion that amounts to a falsification of the report. He protests and in effect is told that his job is at stake.

Jolly, *supra* note 21, at 114.

> A young engineer testifies that he was "asked to present 'edited' results of a reliability study; I refused, and nearly got fired. I refused to defraud the customer . . . ."

Baumhart, *supra* note 19, at 166. The ethical dilemma of the employed engineer is gaining significance because "the practice of engineering is increasingly a corporate endeavor carried out by employees; the engineers in private practice are a dwindling portion of the profession." Labine, *Where is Engineering Pointing?*, CHEMICAL ENGINEERING, Oct. 26, 1964, at 138.

23. *Cf.* Maddock, *The Corporation Law Department*, HARV. BUS. REV., March-April 1952, at 110, 123:

> The thought has been expressed that really good lawyers will refuse to be placed in the position of being subject to the whims of business executives or company politics. The facts do not support this conclusion, at least in the case of most of the corporations with which I am familiar.

It has been estimated that more than 25,000 lawyers practice in corporate law departments. A.B.A., LAW PRACTICE IN A CORPORATE LAW DEPARTMENT 1 (1964).

24. *See* Baumhart, *supra* note 19, at 164-5:

> As controller, I prepared a P&L statement which showed a loss. An executive vice president tried to force me to show a profit in order to present it to a bank for a line of credit. I refused and I was fired on the spot.

25. In Odell v. Humble Oil & Refining Co., 201 F.2d 123 (10th Cir.), *cert. denied*, 345 U.S. 941 (1953), the employees were subpoenaed to appear as witnesses, gave testimony detrimental to the employer and were fired as a result. Compare the prohibition on discharge of an employee who testifies in any proceeding under the Fair Labor Standards Act, 29 U.S.C. § 215(a) (3) (1964). *See* Mitchell v. Dyess, 180 F. Supp. 852 (S.D. Ala. 1960); Wirtz v. C. H. Valentine Lumber Co., 236 F. Supp. 616 (E.D. S.C. 1964).

26. Christy v. Petrus, 365 Mo. 1187, 295 S.W.2d 122 (1956) (employee fired for filing workmen's compensation claim against employer).

27. Mitchell v. Stanolind Pipe Line Co., 184 F.2d 837 (10th Cir. 1950) (employee discharged for bringing assault and battery action against fellow employee).

28. *Cf.* United States Fidelity & Guaranty Co. v. Millonas, 206 Ala. 147, 89 So. 732 (1921).

29. In Payne v. Western & A. R.R., 81 Tenn. 507 (1884), *overruled on other grounds*,

How frequent or widespread such abuses are is open to question. Undoubtedly many employers would not think of engaging in such practices. But it is equally certain that some less scrupulous employers are unable to resist exercising a power within their grasp. Any difficulty in assessing the exact extent of the danger may well be attributable to the fact that where the exercise of individual freedom is not guaranteed there is also no freedom to complain.

What is important is that such abuses, however common or uncommon, should not go unremedied. Whether for the sake of providing specific justice for the afflicted individual, deterring a practice which poses an increasingly serious threat to personal freedom generally, or instilling into employers a general consciousness of and respect for the individuality of the employee, the law should confront the problem.

III. Inadequacy of Existing Limitations on the Employer's Right of Discharge

The labor union, of course, has been a significant source of protection for the employee from all sorts of arbitrary action by the employer. Among the many managerial rights which have been limited by collective bargaining agreements is the right of discharge.[30] Through the "just cause" provisions typically found in collective agreements unions not only protect their constituents from discharges for ulterior purposes, but also prohibit discharges for no reason or for reasons erroneously believed by the employer to be justified.[31] But while unions have done much to correct the imbalance between employers and employees, the assumption that they stand as the universal protectors of all employees at every echelon of employment would be an obvious and gross exaggeration. Less than a quarter of the American working population is now covered by collective bargaining agreements.[32] And there is some well-founded speculation that an accelerating displacement of workers by automation is causing union ranks to shrivel.[33] Then, too, there are many

---

Hutton v. Watters, 132 Tenn. 527, 179 S.W. 134 (1915), for example, the employer issued the following order:

> Any employee of this company . . . who trades with L. Payne from this date will be discharged.

See also Vaughn v. State, 36 Ariz. 32, 282 P. 277 (1929); Hackney v. Fordson Coal Co., 230 Ky. 362, 19 S.W.2d 989 (1929).

30. *See generally* M. Gollub, Discharge for Cause (1948); Holly, *The Arbitration of Discharge Cases: A Case Study*, in Critical Issues in Labor Arbitration 1 (J. McKelvey ed. 1957); Note, *Discharge in the "LAW" or Arbitration*, 20 Vand. L. Rev. 81 (1966).

31. During the past ten years job security has been a particularly significant subject of collective bargaining. *See* Business Week, Oct. 17, 1964, at 45:

> The U.S. Steelworkers seemingly evoked a vision of utopia—at least for its own members—when it described "total job security" as the goal to be pursued in future contract negotiations with the basic steel industry.

32. In 1964, the proportion of union members in the total labor force of the United States was 21.9 percent. 89 Monthly Labor Rev. 510 (1966).

33. From 1956 to 1964 the proportion of union members in the labor force slipped from 24.8 to 21.9 percent. 89 Monthly Labor Rev. 510 (1966).

types of employees, like professionals[34] and other members of the white-collar class,[35] whose numbers are increasing with the advances of modern technology,[36] who have generally preferred not to be represented by labor unions. For such employees it is no answer to suggest that they should seek salvation in unions—that in order to maintain their personal autonomy in the face of the huge industrial employer they should surrender it to the massive labor union.[37] In short, a generally satisfactory solution to the problem of the abusive exercise of employer power does not lie in recourse to labor unionism.

Aside from the protection offered by limitations on the right of discharge reached by collective bargaining, just cause limitations on the right are found in individually negotiated contracts of employment for a specific term. It may be asked why the non-union employee could not protect himself from unwarranted invasion of his personal rights by means of such an individually negotiated contract. It is said that the employer's promise to employ for a definite period of time "is not terminable by him 'at will' after the employee has begun or rendered some of the requested service," and that the employee nevertheless "has retained the power and legal privilege of terminating the employment 'at will.' "[38] But the answer here seems fairly obvious—individual employees are simply not in a position to exact such contracts from their employers. Only the unusually valuable employee has sufficient bargaining power

---

The loss of union membership is not a temporary setback pending the organization of white-collar employees and engineers but the earlier stages of a permanent decline.

J. K. GALBRAITH, THE NEW INDUSTRIAL STATE 263 (1967). *See also* Barkin, *The Decline of the Labor Movement*, in THE CORPORATION TAKE-OVER 223-45 (A. Hacker ed. 1964).

34. *See* Goldstein, *Some Aspects of the Nature of Unionism Among Salaried Professionals in Industry*, 20 AMERICAN SOCIOLOGICAL REVIEW 199 (1955). *See also* Strauss, *Professional or Employee-Oriented: Dilemma for Engineering Unions*, 17 IND. & LAB. REL REV. 519 (1964), where the author explores some possible reasons for a decline in unionism among professional engineers. Some have argued that labor unionism and professionalism are simply incompatible. *See, e.g.*, Morse, *Engineering Ethics—From the Viewpoint of Industry*, 45 JOURNAL OF ENGINEERING EDUCATION 214 (1954). The National Society of Professional Engineers has taken that position. Barkin, *supra* note 33, at 241

35. *Cf.* Blum, *Prospects for Organization of White-Collar Workers*, 87 MONTHLY LABOR REV 125 (1964). In 1964 the number of white-collar workers belonging to unions was approximately 2.6 million, or 14.4 percent of total union membership, 89 MO. LAB. REV. 511 (1966).

36. On the increase in numbers of white-collar workers due to automation, see B. KIRSH, AUTOMATION AND COLLECTIVE BARGAINING 20-21 (1964).

37. Having become large and powerful organizations in their own right, unions have also demonstrated a potential for compromising the rights of the people dependent upon them. While the law has imposed no general duty of fair treatment on employers, the duty of fair representation has been developed to afford some elementary due process type of protection against unfair treatment or oppression by a union of its members. *See* Vaca v. Sipes, 386 U.S. 171 (1967); Humphrey v. Moore, 375 U.S. 335 (1964); Ford Motor Co. v. Huffman, 345 U.S. 330 (1953); Tunstall v. Locomotive Firemen, 323 U.S. 210 (1944); Steele v. Louisville & N.R.R., 323 U.S. 192 (1944). *See generally* Cox, *The Duty of Fair Representation*, 2 VILL. L. REV. 151 (1957); Hanslowe, *The Collective Agreement and the Duty of Fair Representation*, 14 LAB. L. J. 1052 (1963); Summers, *Individual Rights in Collective Agreements: A Preliminary Analysis*, 9 BUFF. L. REV. 239 (1960); Summers, *Individual Rights in Collective Agreements and Arbitration*, 37 N.Y.U. L. REV. 362 (1962).

38. 1A CORBIN, CONTRACTS § 152, at 14 (1963).

to obtain a guarantee that he will be discharged during a specified term of employment only for "just cause." It seems fair to estimate that only a very small portion of the non-unionized employees in this country have succeeded in so altering the presumptively at will nature of the employment relationship. And even for the employee who can obtain such a contract, the term must expire sooner or later; thus his exposure to employer coercion, especially at or near the time of expiration, is still great. Individually negotiated limitations on the employer's right of discharge are not a promising means of solving the problem of employer coercion.

A number of state statutes make it a crime for an employer to coerce or attempt to coerce his employees in certain respects. Some of these statutes, for example, make it unlawful for an employer to prevent employees from engaging in political activities[39] or to compel or prohibit an employee's purchasing goods from a particular concern.[40] The trouble with such statutory provisions is that they are criminal prohibitions which provide no specific redress for the injured employee.[41] It is unlikely that many complaints are made against employers who actually violate such laws. An employee, who is the logical one to file the complaint, would have to be extraordinarily courageous to risk loss of his employment, for which the law provides him no means of recompense, in order to secure the slight satisfaction of seeing his employer suffer the statutory penalties. For the employee who cherishes his and his family's livelihood, the better part of valor may be to submit to the employer's improper demand. These statutory provisions are few in number, narrow in scope, and of doubtful effect even where they do apply. But they at least demonstrate some recognition of the idea that the employer's control over his employee should be limited.

An extra-legal limitation, the employer's interest in maintaining a reputation for fair treatment of employees, may also deter invasion of the employee's freedom and integrity.[42] Obviously, an employer who becomes known for unduly compromising the individuality of his employees may find it diffi-

39. *See, e.g.*, ARIZ. REV. STAT. ANN. § 16-1306 (1956); CAL. LABOR CODE §§ 1101-02 (West 1955); MASS. ANN. LAWS ch. 56, § 33 (1964); MINN. STAT. ANN. § 211.24 (1962); MO. ANN. STAT. § 129.080 (1966); WIS. STAT. ANN. § 12.19 (1967).

40. *See, e.g.*, ARIZ. REV. STAT. ANN. § 23-203 (1956); CAL. LABOR CODE § 450 (West 1955); FLA. STAT. ANN. § 448.03 (1966).

41. In some instances employees have attempted to base a civil action on such statutory prohibitions, but it has been held that the statutes create no right to recover damages. *See, e.g.*, Christy v. Petrus, 365 Mo. 1187, 295 S.W.2d 122 (Mo. 1956); Bell v. Faulkner, 75 S.W.2d 612 (Mo. App. 1934). An exception exists where the statutory prohibition is supplemented by a provision which states that "[n]othing . . . shall prevent the injured employee from recovering damages from his employer for injury suffered through a violation . . . ." Lockheed Aircraft Corp. v. Superior Court, 28 Cal. 2d 481, 486, 171 P.2d 21, 25 (1946).

42. Compare Professor Berle's thesis that big corporations will eventually become, albeit in the not-too-immediate future, responsive to the needs and the rights of individuals through the development of a "corporate conscience." A. BERLE, THE 20TH CENTURY CAPITALIST REVOLUTION (1954).

cult to hire and keep them. But, especially during times of an abundant labor supply, an employer may think it unnecessary to be concerned about his reputation in this respect; and even if he is concerned, he may be able to exert his repressive influence to silence those employees who are affected, and thereby assure that his coercion will never become known. The interest in maintaining a favorable reputation cannot be regarded as a very substantial deterrent to the employer who is tempted to bend his employees to his will.

Finally, the employer's interest in retaining the particular employee cannot be regarded as a significant practical limitation on the employer's control. While a few employees might possess talents so unusual and important that the employer would not risk losing them, most employees are not irreplaceable. The great majority of employees realize that they are expendable, and this realization renders them easy prey to the employer's overreaching demands.

IV. The Employee's Need for a Personal Damage Remedy

The existing sources of protection for the employee are patently inadequate. The question arises whether any other kind of sanction might be used. An appropriate legal response would be to confer on the afflicted employee a personal remedy for any damage he suffers when discharged as a result of resisting his employer's attempt to intimidate or coerce him in a way which bears no reasonable relationship to the employment. For convenience, a discharge so motivated might be termed an "abusive" discharge.

The employee faced with the prospect of losing his job can ordinarily anticipate the expenses of searching for new employment, losing earnings in the meanwhile, and perhaps being forced ultimately to settle for less remunerative employment at some distant place. Beyond these economic losses, he may also fear the stigma and mental anguish which normally accompany being fired.[43] It has been pointed out "that white-collar employees, unlike their brothers in blue collars, are psychologically unprepared for the loss of security and status following on unemployment."[44] That is, the fear of discharge, and thus the vulnerability to employer coercion, is especially acute among professional and other white-collar employees—the very ones whose numbers are increasing and whose jobs are least likely to be protected by collective bargaining agreements.[45]

---

43. In Comerford v. International Harvester Co., 235 Ala. 376, 377, 178 So. 894, 895 (1938), for example, the plaintiff-employee alleged that as a consequence of his wrongful discharge he "was greatly humiliated and embarrassed, and was caused to suffer great mental anguish and was caused to lose a lucrative and profitable position, and was put to great trouble, expense, annoyance and inconvenience in and about obtaining new employment, and was caused to be without employment for a long period of time, and to lose a large sum of money for salary he would otherwise have received . . . ."

44. Hacker, *Introduction: Corporate America*, in The Corporation Take-over 9 (A. Hacker ed. 1964).

45. *Cf.* J. K. Galbraith, The New Industrial State 267 (1967):

If the employee had some assurance that he would not have to bear such losses, he would be in a far better position to withstand oppression at the hands of his employer. Such assurance could be provided by arming the employee with a damage action where his discharge is caused by a refusal to submit to the employer's improper or overreaching demands. It should be emphasized that so to limit the employer's right of discharge would not give blanket protection to the employee's interest in job security. There is a distinction between the right to employment and the right of the employee not to be obliged to his employer in ways bearing no legitimate connection to the employment.[46]

Recognition of such a cause of action would of course tend to deter an employer from discharging an employee for an abusive reason. Further, employers would face the danger that a subsequent discharge, even though for good cause, might be associated with a prior attempt of the employer to interfere with the employee's individuality. Thus the fear of lawsuits would have the salutary effect of discouraging improper attempts to interfere with the employee's freedom or integrity, even when the employer does not intend to discharge the employee for refusing to submit to his desires. But beyond the more immediate effects, and perhaps more important, legal protection for the abusively discharged employee would inevitably develop a keener awareness of and greater respect for the individuality of the employee. Indeed, it is conceivable that this consciousness of employee individuality would do much toward solving the dilemma of the organization man, the employee in middle management whom William H. Whyte saw as too prone to identify with his employer and especially subject to the power of his employing organization.[47]

The remedy suggested here for the abusively discharged employee is not without parallel. The National Labor Relations Board is empowered to grant damages (in the form of "back pay"), and reinstatement as well, to an employee whose discharge is found to have stemmed from his involvement in certain labor union activities.[48] Similar remedies are granted by the various Civil Rights and Fair Employment Practices Commissions which have been given the responsibility of protecting against discriminatory practices in both

---

> [M]odern technology opens the way for a massive shift from workers who are within the reach of unions to those who are not.

See also notes 34 to 36 *supra* and accompanying text.

46. Compare the distinction drawn with respect to public employment in Dotson, *The Emerging Doctrine of Privilege in Public Employment*, 15 PUB. AD. REV. 77, 87 (1955) (emphasis in original):

> Even if it were granted that no constitutional right to employment could be established, this concession would not imply that, by virtue of public employment, an individual might be deprived of his *other* constitutional rights.

47. W. H. WHYTE, THE ORGANIZATION MAN (1956). Judicial sanctions could not, however, cure this problem to the extent that it stems not from the coercive power of the employer but from the employee's own psychological need to identify his own goals with those of his employer.

48. National Labor Relations Act § 10(c), 29 U.S.C. § 160(c) (1964).

hiring and firing.[49] And judicial and quasi-judicial remedies are available to wrongfully discharged employees in a number of continental European countries.[50]

The Federal Automobile Dealer Franchise Act of 1956[51] also provides a precedent, in the damage remedy it gives to wrongfully disenfranchised dealers, and in its economic background.[52] The expendability of the automobile dealer introduced an imbalance in bargaining power into his relationship with the manufacturer. As with the employer's power over the employee, the real source of the manufacturer's ability to take unfair advantage of the dealer resided in its power to terminate the relationship.[53] The courts failed to respond to the dealer's need for protection,[54] but the dealer's strong lobby[55] brought about a congressional investigation[56] which uncovered a number of abusive uses of the manufacturer's power. The upshot was the 1956 Act with its explicitly stated purpose "to guarantee the one party freedom from coercion, intimidation, or threats of coercion from the other party"[57]—a goal identical with that toward which this entire discussion is directed.

---

49. *See* note 127 *infra*.

50. *See generally* THE EMPLOYMENT RELATION AND THE LAW 806-09 (B. Aaron & R. Mathews eds. 1957).

51. 15 U.S.C. §§ 1221-25 (1964).

52. *See generally* Kessler, *Automobile Dealer Franchises: Vertical Integration by Contract*, 66 YALE L.J. 1135 (1957); Brown & Conwill, *Automobile Manufacturer-Dealer Legislation*, 57 COLUM. L. REV. 219 (1957).

53. The manufacturer is . . . in a position to cause serious financial loss to the dealer through use of its power to terminate the relationship. By threatening to terminate the relationship the manufacturer can force the dealer to make unwanted purchases or to take other disadvantageous steps in the conduct of his business.

Brown & Conwill, *supra* note 52, at 222-23.

The real source of the manufacturer's power over the dealer lies in the termination provisions of the franchise agreements.

Strand & French, *The Automobile Dealer Franchise Act: Another Experiment in Federal Class Legislation*, 25 GEO. WASH. L. REV. 667, 668 (1957).

54. The usual ground given by the courts in refusing to enforce the automobile dealer franchise agreements was a lack of "mutuality." *See, e.g.*, Ford Motor Co. v. Kirkmeyer Motor Co., 65 F.2d 1001 (4th Cir. 1933); Oakland Motor Car Co. v. Indiana Auto. Co., 201 F. 499 (7th Cir. 1912); Superior Motor Co. v. Chevrolet Motor Co., 112 Kan. 522, 212 P. 100 (1923). Compare the discussion accompanying notes 67 to 71 *infra* of the use by the courts of the requirement of mutuality of obligation in refusing to enforce employment contracts.

55. The dealers spoke chiefly through the National Association of Automobile Dealers. Kessler, *supra* note 53, at 1167; Brown & Conwill, *supra* note 52, at 223.

56. *See Hearings on H.R. 11360 and S. 3879 Before the Antitrust Subcomm. No. 5 of the House Comm. on the Judiciary*, 84th Cong., 2d Sess., ser. 26 (1956); *Hearings Before the Subcomm. on Automobile Marketing Practices of the Senate Committee on Interstate and Foreign Commerce*, 84th Cong., 2d Sess. (1956).

57. 15 U.S.C. § 1221(e) (1964). The fact that there have been few reported cases brought under this Act in which dealers have prevailed has caused a number of commentators to leap to the conclusion that the legislation has been ineffective. *See, e.g.*, Freed, *A Study of Dealers' Suits Under the Automobile Dealers' Franchise Act*, 41 U. DET. L.J. 245, 256-61 (1964); Comment, *The Automobile Dealer Franchise Act of 1956—An Evaluation*, 48 CORNELL L.Q. 711, 741-42 (1963). But in the most thorough-going study ever made of this legislation and its effects it has been pointed out that these commentators overlook the benefit which the legislation has produced in acting as a spur to the creation of private machinery for the settlement of manufacturer-dealer termination disputes. *See* Macauley, *Changing a Continuing Relationship Between a Large Corpora-*

V. The Traditional Rule and Its Foundations in Policy

Recognizing a cause of action for the abusively discharged employee would, of course, significantly limit the traditional absolute right of discharge. Thus it seems appropriate to take a closer look at this traditional right to see if there is anything behind it in policy or theory which ought to prevent the imposition of such a limitation.

As is often done with timeworn rules which cannot otherwise be justified, the "arbitrary right of the employer to employ or discharge labor, with or without regard to actuating motives," is a proposition "settled beyond peradventure."[58] It is likely that the rule became "settled beyond peradventure" at a time when the words "master" and "servant" were taken more literally than they are now and when, as in early Roman law, the rights of the servant, like the rights of any other member of the household, were not his own but those of his *pater-familias*.[59] Overtones of this ancient doctrine are clearly discernible in a late nineteenth century opinion which rationalized the employer's absolute right of discharge as follows:

> May I not refuse to trade with any one? May I not forbid my family to trade with any one? May I not dismiss my domestic servant for dealing, or even visiting, when I forbid? And if my domestic, why not my farm-hand, or my mechanic, or teamster?[60]

Such a philosophy of the employer's dominion over his employee may have fit the rustic simplicity of the days when the farmer or small entrepreneur, who may or may not have employed others, was the epitome of American individualism. But the philosophy is incompatible with these days of large, impersonal, corporate employers; it does not comport with the need to preserve individual freedom in today's job-oriented, industrial society. Nevertheless, the courts have strictly adhered to the traditional rule that in the absence of a statute or agreement specifically limiting the right of discharge, the employer may discharge his employee at any time for any reason.

A brief survey of the constitutional history of the traditional rule will serve further to demonstrate that the philosophical underpinnings of the rule have fallen into decay. In two early twentieth century cases, the United States Supreme Court elevated the employer's absolute right of discharge to a constitutionally protected property right. Due process was held to be violated by

---

tion and Those Who Deal With It: Automobile Manufacturers, Their Dealers, and the Legal System (pts. I & II), 1965 Wis. L. Rev. 483, 740 at 857. "Their mistaken appraisal rests on their failure to consider the operation of the less formal parts of the legal system and the private systems of planning and adjustment which were created in response to, and are supported by, the legal system."

58. Union Labor Hosp. Ass'n v. Vance Redwood Lumber Co., 158 Cal. 551, 555, 112 P. 886, 888 (1910).

59. *See* Sayre, *Inducing Breach of Contract*, 36 Harv. L. Rev. 663 (1923).

60. Payne v. Western & A.R.R., 81 Tenn. 507, 518 (1884), *overruled on other grounds*, Hutton v. Watters, 132 Tenn. 527, 179 S.W. 134 (1915).

any law which interfered with "the right of the purchaser of labor to prescribe the conditions upon which he will accept such labor from the person offering to sell it."[61] In both cases the employee lost his job not because of the way he performed his work but because of his membership in a labor union. And in both cases, the Supreme Court struck down anti-yellow-dog-contract legislation which made it a crime for an employer to discharge an employee for such a reason.

In the first case, *Adair v. United States*,[62] the Court stated the general rule that "the right of the employee to quit the service of the employer, for whatever reason, is the same as the right of the employer, for whatever reason, to dispense with the services of such employee." The Court then declared that federal legislation which disturbs such equality by compelling one person to retain the personal services of another is "an invasion of the personal liberty as well as of the right of property guaranteed by the 5th Amendment."

A few years later, similar state legislation was invalidated under the due process clause of the fourteenth amendment. In *Coppage v. Kansas* the Court elaborated on its reasons for concluding that the employer's right to hire and fire whom he wishes for any or no reason was a constitutionally protected property right.[63] Justice Pitney's majority opinion clearly displayed a persisting belief in inviolable rights of property and uninhibited freedom of contract:

> As to the interest of the employees, it is said by the Kansas Supreme Court . . . to be a matter of common knowledge that "employees, as a rule, are not financially able to be as independent in making contracts for the sale of their labor as are employers in making contracts of purchase thereof." No doubt, wherever the right of private property exists, there must and will be inequalities of fortune; and thus it naturally happens that parties negotiating about a contract are not equally unhampered by circumstances . . . . [S]ince it is self-evident that, unless all things are held in common, some persons must have more property than others, it is from the nature of things impossible to uphold freedom of contract and the right of private property without at the same time recognizing as legitimate those inequalities of fortune that are the necessary result of the exercise of those rights.[64]

Justice Day, dissenting, did not subscribe to the *laissez faire* philosophy expressed in the majority opinion. He was of the view that laws could be enacted to prevent the undue or oppressive exercise of authority by employers in making contracts with employees:

> It may be that an employer may be of the opinion that membership of his employees in the National Guard, by enlistment in the militia

61. Coppage v. Kansas, 236 U.S. 1, 10 (1915).
62. 208 U.S. 161 (1908).
63. 236 U.S. 1 (1915).
64. *Id.* at 17.

> of the State, may be detrimental to his business. Can it be successfully contended that the State may not, in the public interest, prohibit an agreement to forego such enlistment as against public policy? Would it be beyond a legitimate exercise of the police power to provide that an employee should not be required to agree, as a condition of employment, to forego affiliation with a particular political party, or the support of a particular candidate for office? It seems to me that these questions answer themselves.[65]

Justice Day also expressed the desirability of infusing an element of equality into a relationship characterized by imbalance:

> I think that the act now under consideration, and kindred ones, are intended to promote the same liberty of action for the employee, as the employer confessedly enjoys. The law should be as zealous to protect the constitutional liberty of the employee as it is to guard that of the employer. A principal object of this statute is to protect the liberty of the citizen to make such lawful affiliations as he may desire with organizations of his choice. It should not be necessary to the protection of the liberty of one citizen that the same right in another citizen be abridged or destroyed.[66]

The philosophy articulated by Justice Day was destined to gain the ascendancy, and *Adair* and *Coppage* to be sapped of their authority. In upholding the National Labor Relations Act in *NLRB v. Jones & Laughlin Steel Corp.*,[67] the Supreme Court expressly approved the Act's protection of the right of employees to unionize free of intimidation and coercion by employers. In so holding the Court made the especially noteworthy observation that the N.L.R.A. did not interfere with the "normal" exercise of the right of discharge, but was aimed only at prohibiting employers from "using the right of discharge as a means of intimidation and coercion."[68]

The demise of *Adair* and *Coppage* demonstrates that from the standpoint of sound social policy, and thus as a matter of constitutional principle, the traditional rule can no longer be justified. The industrial revolution made an anachronism of the absolute right of discharge by destroying the classical ideal of complete freedom of contract upon which it is based.[69] It is obvious, moreover, that the idea expressed in *Jones & Laughlin* of not interfering with the normal exercise of the right of discharge, but preventing employers from using it as a means of oppression, is worthy of the most general application. If the principle of collective bargaining justified placing such a limitation on the employer because the "employee is sensitive and responsive to even the

65. *Id.* at 37.
66. *Id.* at 40.
67. 301 U.S. 1 (1937). *See also* Associated Press v. NLRB, 301 U.S. 103 (1937); Texas & N.O.R.R. v. Brotherhood of Railway Clerks, 281 U.S. 548 (1930).
68. 301 U.S. at 45-46.
69. "The system of 'free' contract described by nineteenth century theory is now coming to be recognized as a world of fantasy, too orderly, too neatly contrived and too harmonious to correspond with reality." Dawson, *Economic Duress and the Fair Exchange in French and German Law*, 11 TUL. L. REV. 345 (1937).

most subtle expression on the part of the employer, whose good will is so necessary,"[70] the same reason would seem to warrant broad protection of the right of all employees to function as individuals in a free society.

VI. The Traditional Rule and the Law of Contracts

It appears that the traditional rule has survived largely because of the sustenance it has received from the law of contracts. From the contractual principle of mutuality of obligation, it has been reasoned that if the employee can quit his job at will, then so, too, must the employer have the right to terminate the relationship for any or no reason.[71] Indeed, there is a multitude of cases in which even contracts for permanent employment, that is, for indefinite terms, have been held unenforceable on the ground that they lack mutuality of obligation.[72] But these cases demonstrate that mutuality is a high-sounding phrase of little use as an analytical tool.[73] If the employee in addition to his services has given other "good" consideration, such as foregoing a claim against the employer or giving up a business to accept the employment, the agreement will be enforced on behalf of the employee even though he is free to quit at any time.[74] Thus it seems clear that mutuality of obligation is not an inexorable requirement and that lack of mutuality is simply, as many courts have come to recognize,[75] an imperfect way of referring to the real

---

70. NLRB v. Griswold Mfg. Co., 106 F.2d 713, 722 (3d Cir. 1939).

71. It is ironic that application of the mutuality notion to the employment relationship has been expressed as arising out of a primary concern for the freedom of the employees:

> An employee is never presumed to engage his services permanently, thereby cutting himself off from all chances of improving his condition; indeed, in this land of opportunity it would be against public policy and the spirit of our institutions that any man should thus handicap himself; and the law will presume . . . that he did not so intend. And if the contract of employment be not binding on the employee for the whole term of such employment, then it cannot be binding upon the employer; there would be lack of "mutuality."

Pitcher v. United Oil & Gas Syndicate, Inc., 174 La. 66, 69, 139 So. 760, 761 (1932).

72. *See, e.g.*, Meadows v. Radio Indus., 222 F.2d 347 (7th Cir. 1955); Lord v. Goldberg, 81 Cal. 596, 22 P. 1126 (1889); Hope v. National Airlines, Inc., 99 So. 2d 244 (Fla. App.), *cert. denied*, 102 So. 2d 728 (Fla. 1958); Rape v. Mobile & O.R.R., 136 Miss. 38, 100 So. 585 (1924).

73. Corbin, for one, has pointed out that there is "vagueness and inconsistency and error" in use of the phrase "mutuality." 1A Corbin, Contracts § 152 (1963).

74. *See, e.g.*, F. S. Royster Guano Co. v. Hall, 68 F.2d 533 (4th Cir. 1934); Alabama Mills v. Smith, 237 Ala. 296, 186 So. 699 (1939); Seifert v. Arnold Bros., Inc., 138 Cal. App. 324, 31 P.2d 1059 (1934); Edwards v. Kentucky Util. Co., 286 Ky. 341, 150 S.W.2d 916 (1941) (dictum); Yellow Poplar Lumber Co. v. Rule, 106 Ky. 455, 50 S.W. 685 (1899); Kirkley v. F. H. Roberts Co., 268 Mass. 246, 167 N.E. 289 (1929).

> If the employee gives a sufficient consideration for the employer's promise, the lack of mutuality of obligation is immaterial.

3A Corbin, Contracts § 684, at 229 (1960).

75. *See, e.g.*, Armstrong Paint & Varnish Works v. Continental Can Co., 301 Ill. 102, 108, 133 N.E. 711, 714 (1921):

> While consideration is essential to the validity of a contract, mutuality of obligation is not. Where there is no other consideration for a contract, the mutual promises of the parties constitute the consideration, and these promises must be binding on both parties or the contract falls for want of consideration, but, where there is any other consideration for the contract, mutuality of obligation is not

obstacle to enforcing any kind of contractual limitation on the employer's right of discharge—lack of consideration.

It might plausibly be argued that the requisite consideration is to be found in the employee's past service. Indeed, there seems to be some truth in the assertion that an employee who spends a significant part of his working life working for one employer to the exclusion of others has conferred a substantial benefit on the employer. There seems to be even more truth in the assertion that such an employee has suffered a real detriment in the irretrievable loss of productive years, especially when his seniority and experience are not likely to be readily transferable to new employment. These arguments are stronger when the employee is to receive death or retirement benefits which have further induced him to remain in the employer's service.[76] It is apparent, however, that the courts regard the employee as fully recompensed for his services by wages; nothing is left over to support any promise of continued employment.[77] So long as courts are unwilling to liberalize their notions of the consideration necessary to support promise of permanent employment, there is little reason to expect that they will do so in order to uphold a promise not to discharge an at will employee for ulterior reasons,[78] an implied promise for which there is at best an uncertain factual basis.[79]

If there is anything in contract law which seems likely to advance the present inquiry it is the growing tendency to protect individuals from contracts of adhesion—from overreaching terms often found in standard form

---

essential. If mutuality, in a broad sense, were held to be an essential element in every valid contract to the extent that both contracting parties could sue on it, there could be no such thing as a valid unilateral or option contract.

Similar statements can be found in Meurer Steel Barrel Co. v. Martin, 1 F.2d 687, 688 (3d Cir. 1924); J. C. Millett Co. v. Park & Tilford Distillers Corp., 123 F. Supp. 484, 493 (N.D. Cal. 1954); Levin v. Perkins, 12 Wis. 2d 398, 403, 107 N.W.2d 492, 495 (1961).

76. The combined presence of inducement and detrimental reliance strongly suggests the possible application of section 90 of the RESTATEMENT (SECOND) OF CONTRACTS (Tent. Draft No. 2 (1965) (promissory estoppel). *But see* MacCabe v. Consolidated Edison Co., 30 N.Y.S.2d 445, 447 (N.Y. City Ct. 1941), where the court said, in considering the employee's rights under a retirement plan:

> "In this state, the rule is settled that, unless a definite period of service is specified in the contract, the hiring is at will; and the master has the right to discharge and the servant to leave at any time." . . . If the hiring can be terminated at will, and provision for an employee's retirement is not embodied in a juridically recognizable obligation of the employer, then whatever provision may be made is, in the eye of the law, not a right but a gift.

77. *See* Note, *Employment Contracts of Unspecified Duration*, 42 COLUM. L. REV. 107 (1942).

78. *But cf.* St. Louis, B. & M. Ry. v. Booker, 5 S.W.2d 856, 859 (Tex. Civ. App. 1928), *cert. denied*, 279 U.S. 52 (1929), where an express promise not to discharge at will employees "unjustly" was enforced on the ground that it was "part of the consideration inducing employees to enter and remain in the service, and the continued performance of the duties of their employment is a valuable consideration to the railway company."

79. But compare the following observation by Judge Clark in Parev Prods. Co. v. I. Rokeach & Sons, 124 F.2d 147, 149 (2d Cir. 1941):

> When we turn to the precedents we are met at once with the confusion of statement whether a covenant can be implied only if it was clearly "intended" by the parties, or whether such a covenant can rest on principles of equity. . . . One may perhaps conclude that in large measure this confusion arises out of the reluctance of courts to admit that they were to considerable extent "remaking" a contract in situations where it seemed necessary and appropriate so to do.

contracts used by large commercial establishments.[80] Judicial disfavor of contracts of adhesion has been said to reflect the assumed need "to protect the weaker contracting party against the harshness of the common law and what they think are abuses of freedom of contract."[81] The same philosophy seems to provide an appropriate answer to the argument, which seems still to have some vitality,[82] that "the servant cannot complain, as he takes the employment on the terms which are offered him."[83]

The adhesion concept has been applied to employment contracts—to invalidate agreements exempting the employer from liability for negligent injury to the employee.[84] But in view of the fact that the doctrine is apparently used only to invalidate overreaching contractual terms and not to raise obligations where none previously existed, it would be unrealistic to suggest that it could be a basis for limiting the employer's right of discharge. Even if an implied promise not to discharge the employee for an abusive reason could be built on the adhesion concept, such a promise would still lack supporting consideration. Inability to apply the doctrine directly to judicial protection of employee individuality, however, does not weaken its force as a philosophical analogue.[85]

In the last analysis, then, it is not policy but the technical difficulty of relaxing the rather rigid rules of consideration which makes it unlikely that the employer's right to terminate the at will employment relationship can be limited under contract law.

VII. The Traditional Rule and the Law of Torts: The Source of a Possible Judicial Limitation on the Employer's Right to Terminate the At Will Employment Relationship

If a damage remedy were to be extended to abusively discharged employees, it would protect personality interests of the employee—interests which by definition have no legitimate connection with the employment relationship

80. *See, e.g.*, Campbell Soup Co. v. Wentz, 172 F.2d 80 (3d Cir. 1948); Fairfax Gas & Supply Co. v. Hadary, 151 F.2d 939 (4th Cir. 1945); Thomas v. First Nat'l Bank, 376 Pa. 181, 101 A.2d 910 (1954); Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960). See also § 2-302(1) of the Uniform Commercial Code, relating to contracts for the sale of goods:

> If the court as a matter of law finds the contract or any clause of the contract to have been unconscionable at the time it was made the court may refuse to enforce the contract, or it may enforce the remainder of the contract without the unconscionable clause, or it may so limit the application of any unconscionable clause as to avoid any unconscionable result.

81. Kessler, *Contracts of Adhesion—Some Thoughts About Freedom of Contract*, 43 Colum. L. Rev. 629, 636 (1943).

82. *See* Garrity v. New Jersey, 385 U.S. 493, 499 (1967).

83. Justice Holmes in McAuliffe v. New Bedford, 155 Mass. 216, 29 N.E. 517 (1892).

84. *E.g.*, Blanco v. Phoenix Compania De Navegacion, S.A., 304 F.2d 13 (4th Cir. 1962); Johnston v. Fargo, 184 N.Y. 379, 77 N.E. 388 (1906).

85. Like the adhesion concept, the doctrine of economic duress, as one of the legal concepts used to alleviate some of the harsh effects of unequal economic power, bears some analogy here. But the doctrine has apparently been used only to provide restitutionary relief and not to compensate for any loss of prospective advantage. *See* Dawson, *Economic Duress—An Essay in Perspective*, 45 Mich. L. Rev. 253, 282-83 (1947). Hence, chances for application of doctrine in cases of abusive discharge are probably slight.

or "contract." If the employer invades legally protected rights of the employee, for instance by the infliction of bodily injury or by defamation, the existence of the employment "contract" does not stand in the way of determining the employee's rights under the law of torts.[86] Since analogous interests of the employee are at stake when the employer unreasonably attempts to interfere with his personal freedom, it seems reasonable to bypass the law of contracts and its unyielding requirement of consideration by turning to the more elastic principles of tort law for a suitable basis upon which to predicate the discharged employee's action for damages. Such a basis may exist in the various types of tort liability which have evolved in connection with the exercise of a right for an improper, ulterior purpose.

Legal recognition that one can be held liable solely on account of wrongful motives is more or less a twentieth century development.[87] Earlier, the law's general attitude toward bad motives found expression in the oft-stated maxim that "malicious motives make a bad act worse, but they cannot make that wrong which in its own essence is lawful."[88] This maxim, however, has come to be recognized as mere question begging,[89] and there are now many instances of tort liability based largely on the defendant's bad motives. Numerous, for example, are cases in which motive has resulted in liability for building spite fences,[90] for drilling wells to cut off the plaintiff's underground water,[91] for promoting numerous other nuisances designed only to harass the plaintiff,[92] or for unwarranted interference with favorable contractual relationships.[93]

86. *See, e.g.*, Imre v. Riegel Paper Corp., 24 N.J. 438, 132 A.2d 505 (1957), where the employee recovered medical expenses and loss of wages incurred as a result of the negligence of his employer. The employer's liability for negligence has, of course, been largely superseded by the workmen's compensation acts.

87. *See generally* Lewis, *Should the Motive of the Defendant Affect the Question of His Liability*, 5 COLUM. L. REV. 107 (1905); Walton, *Motive as an Element in Torts in the Common and in the Civil Law*, 22 HARV. L. REV. 501 (1909).

88. *See* Jenkins v. Fowler, 24 Pa. 308, 310 (1855). The statement was endorsed in T. COOLEY, A TREATISE ON THE LAW OF TORTS 497 (1st ed. 1879).

89. "This of course merely begs the question, since unless motive is to be eliminated altogether, it must be taken into account in determining whether the act is 'in its essence lawful' in the first place." W. PROSSER, HANDBOOK OF THE LAW OF TORTS 24 (3d ed. 1964).

90. *E.g.*, Larkin v. Tsavaris, 85 So. 2d 731 (Fla. 1956); Hornsley v. Smith, 191 Ga. 491, 13 S.E.2d 20 (1941); Flaherty v. Moran, 81 Mich. 52, 45 N.W. 381 (1890); Hibbard v. Halliday, 58 Okla. 244, 158 P. 1158 (1916); Racich v. Mastiovich, 65 S.D. 321, 273 N.W. 660 (1937); Erickson v. Hudson, 70 Wyo. 317, 249 P.2d 523 (1952).

It has been said that maintenance of a spite fence should be enjoined upon proof "not only that the structure complained of is entirely useless to the [defendant], and without value to her property, but also that it was maliciously erected for the purpose of injuring complainant in the use and enjoyment of his property." Norton v. Randolph, 176 Ala. 381, 390, 58 So. 283, 286 (1912).

91. *E.g.*, Katz v. Walkinshaw, 141 Cal. 116, 74 P. 766 (1903); Gagnon v. French Lick Springs Hotel Co., 163 Ind. 687, 72 N.E. 849 (1904); Barclay v. Abraham, 121 Iowa 619, 96 N.W. 1080 (1903); Chesley v. King, 74 Me. 164 (1882); Stillwater Water Co. v. Farmer, 89 Minn. 58, 93 N.W. 907 (1903).

92. *See generally* Ames, *How Far an Act May Be a Tort Because of Wrongful Motive of the Actor*, 18 HARV. L. REV. 411 (1905); Fridman, *Motive in the English Law of Nuisance*, 40 VA. L. REV. 583 (1954).

93. *See generally* Carpenter, *Interference with Contract Relations*, 41 HARV. L. REV. 728 (1928); Green, *Relational Interests*, 29 ILL. L. REV. 1041 (1935); Sayre, *Inducing*

Many of the cases of tort liability based on the defendant's bad motives have been grouped under the rubric of "prima facie tort," which found expression in the classic statement:

> Now intentionally to do that which is calculated in the ordinary course of events to damage, and which does, in fact, damage another in that other person's property or trade, is actionable if done without just cause or excuse.[94]

It has been pointed out that this is "no more than a form of words emphasizing the importance of motive."[95] But regardless of whether this notion is stated in terms of prima facie tort or bad motives, it lends general support to the proposition that tort liability ought to attach in cases of abusive discharge.[96]

An analogy which seems particularly suited to the case of a discharge caused by improper motives is the tort action designed to prevent the perversion of legal procedures to ulterior purposes—abuse of process.[97] Essential to recovery for abuse of process is the defendant's intent to exercise a right, lawful and valuable in itself, for a purpose other than that for which it was designed. Unlike the related tort of malicious prosecution, the action for abuse of process does not depend on the absence of probable cause—the gist of the action is exercise of a right for an ulterior purpose regardless of whether it can or cannot be otherwise justified.[98] This emphasis on state of mind makes the doctrinal framework of abuse of legal procedures especially suitable for an approach to the problem of abusive dismissals. To elaborate: in *Grainger v. Hill*,[99] the landmark case on abuse of process, the defendants were held liable chiefly because, in the words of one of the judges, "The process was enforced for an ulterior purpose; to obtain property by duress to which the Defendants had no right."[100] Liability should similarly be visited upon the employer who uses his power of discharge for an ulterior purpose and as a means of duress.

Breach of Contract, 36 HARV. L. REV. 663 (1923); Harper, *Interference with Contractual Relations*, 47 NW. U.L. REV. 873 (1953); W. PROSSER, HANDBOOK OF THE LAW OF TORTS § 123 (3d ed. 1964).

94. Lord Bowen in Mogul Steamship Co. v. McGregor, Gow & Co., 23 Q.B.D. 598, 613 (C.A. 1889). For general discussions of the prima facie tort concept, *see* Holmes, *Privilege, Malice and Intent*, 8 HARV. L. REV. 1 (1894); Brown, *The Rise and Threatened Demise of the Prima Facie Tort Principle*, 54 NW. U.L. REV. 563 (1959); Note, *The Prima Facie Tort Doctrine*, 52 COLUM. L. REV. 503 (1952).

95. W. PROSSER, HANDBOOK OF THE LAW OF TORTS § 26 (3d ed. 1964).

96. *Cf.* 51 COLUM. L. REV. 398 (1951).

97. *See generally* RESTATEMENT OF TORTS § 682 (1938); W. PROSSER, HANDBOOK OF THE LAW OF TORTS § 115 (3d ed. 1964); 32 MINN. L. REV. 805 (1948).

98. The difference between malicious prosecution and abuse of process is discussed in Baldwin v. Davis, 188 Ga. 587, 4 S.E.2d 458 (1939).

99. 4 Bing., N.C. 212, 132 Eng. Rep. 769 (C.P. 1838). In *Grainger*, the plaintiff mortgaged his vessel to the defendants under terms giving the plaintiff the right to retain possession. Thereafter, however, the defendants demanded that the plaintiff give up the register to the vessel, without which he could not go to sea. Upon plaintiff's refusal, the defendants obtained his arrest and imprisonment under a writ of capias—in order to compel him to give up something to which the defendants were not entitled under the terms of the mortgage.

100. *Id.* at 681, 132 Eng. Rep. at 774.

It can be said that discharge of the at will employee, like resort to legal procedures, is in its essence lawful. It can also be said, as with the right to invoke the processes of the law, that the employer's right of discharge is too valuable a right to be encumbered with unnecessary limitations. But, as with any individual's right to bring legal action, the law should not allow the employer to exercise his right of discharge in order to effectuate a purpose ulterior to that for which the right was designed.[101] Just as the use of legal processes as a means of extortion gives rise to a damage remedy, so too should the oppressive use of the right of discharge.

While cases dealing with abuse of process provide the most suitable analogy, further support for the employee's action can be derived from the growing body of cases in which interference by a third party with the employee's interest in the at will employment relationship has given rise to recovery in tort.[102] *United States Fidelity & Guarantee Co. v. Millonas*[103] is illustrative. The plaintiff-employee filed a claim with the defendant, the employer's insurer, for compensation for an injury he had suffered in the course of his employment. The defendant told the plaintiff that he would lose his job if he persisted in his claim. Then, by threatening cancellation of the employer's policy, the defendant procured the plaintiff's discharge. The defendant argued that no cause of action could arise from its threatening to do what it had a legal right to do to cancel the employer's policy. The court rejected this argument with the statement that the plaintiff's discharge was not procured as

---

101. The sort of emphasis which our common law has placed on the defendant's bad motives in certain instances may also be found in the civil law concept of liability for "abusive use of a right" or abus de droit. This concept is based on the notion that rights are limited not only in their extent, but also in their exercise, and that there is therefore an abuse of the right if it is exercised with the intention of injuring another. French law has long employed this reasoning in cases where legal processes are used for ulterior purposes and in cases of discharge from employment without just cause. *See generally* Walton, *supra* note 87; Gutteridge, *Abuse of Rights*, 5 CAMB. L.J. 22 (1933).

102. *See, e.g.*, Canuel v. Oskoian, 184 F. Supp. 70 (D.R.I. 1960); London Guarantee & Accident Co. v. Horn, 206 Ill. 493, 69 N.E. 526 (1903); DeMinico v. Craig, 207 Mass. 593, 94 N.E. 317 (1911); Carneso v. St. Paul Union Stockyards Co., 164 Minn. 457, 205 N.W. 630 (1925); Warschauser v. Brooklyn Furniture Co., 159 App. Div. 81, 144 N.Y. Supp. 257 (2d Dep't 1913); Jones v. Leslie, 61 Wash. 107, 112 P. 81 (1910); Mendelson v. Blatz Brewing Co., 9 Wis. 2d 487, 101 N.W.2d 805 (1960); Giblan v. National Amalgamated Labourers' Union, [1903] 2 K.B. 600 (C.A.). The source of common law protection of the employment relationship appears to be the Ordinance and Statute of Labourers, 23 Edw. III, St. 1 (1349) and 25 Edw. III, St. 1 (1350), which were enacted after the Black Death had reduced the labor force in England by almost one-half. 2 W. HOLDSWORTH, A HISTORY OF ENGLISH LAW 459-60 (3d ed. 1923). It was the employer's interest which these fourteenth century statutes sought to protect. The connection between these ancient provisions and modern day protection of the employment relationship, and indeed of all contractual relations, is shown in the landmark decision of Lumley v. Gye, 2 E. & B. 216, 118 Eng. Rep. 749 (Q.B. 1853). In holding the defendant liable for enticing a famous opera singer from the employ of the plaintiff, the court reasoned that even though the relationship of the opera singer and the plaintiff was not of the sort defined in the Statute of Labourers, the injury suffered by the plaintiff was so closely analogous to that encompassed by the Statute that an extension of its underlying principle was warranted.

103. 206 Ala. 147, 89 So. 732 (1921). A case similar to *Millonas* is American Surety Co. v. Schottenbauer, 257 F.2d 6 (8th Cir. 1958).

"the consequence of the exercise of a lawful right, but of the unlawful use of that lawful right by the defendant."[104]

This decision and others like it go beyond merely exemplifying further the role of wrongful motives in tort liability. They find liability not only when the defendant acts from ulterior motives, but also when he acts "without justification."[105] Thus, a defendant may be held liable though he lacks the state of mind which makes the employer's discharge "abusive" in the definition urged here. But despite the imperfection of the analogy, the interference cases demonstrate that an employee's interest in an at will employment relationship is considered deserving of legal protection. And if the interest is protectable, it is difficult to see why it should matter whether the employee's discharge is procured by the abuse of a "lawful right" by a third party or by the employer.

Many of the cases involving interference by a third party with the at will employment relationship concern protection of the employer's valuable interest in keeping his employee.[106] So, it might be argued, if the employee's interest in the relationship is to be protected from interference by the employer as well as by third parties, the employer's interest in preserving the same relationship should be similarly protected from employees. The answer to this argument is that an equation of the rights of the employer with the rights of the employee is inconsistent with the basic inequality in the positions of the two. Moreover the freedom of the individual not to work for a particular employer has come to be regarded as a more valuable right than the freedom of the employer to select his employees. The employer's freedom has been limited by our child labor laws, the legally protected status of labor unions, and the growing body of laws prohibiting discrimination in hiring. And while the employer's right to select his employees has been thus limited, the freedom of the individual to refuse to work has, perhaps because of our abhorrence of slavery, become virtually inviolable. As one court put it in holding a third party liable for procuring the discharge of an employee:

> The right to dispose of one's labor as he will . . . is incident to the freedom of the individual, which lies at the foundation of the government in all countries that maintain principles of civil liberty.[107]

In short, the employee's right to work for whom he chooses is too valuable to be circumscribed or limited to prevent abuse of the almost negligible coercive

104. 206 Ala. at 151, 89 So. at 735.

105. *See* Carpenter, *supra* note 93, at 734-35.

106. It was the employer's interest in the employment relationship which the law first sought to protect. See note 102 *supra.*

107. Berry v. Donovan, 188 Mass. 353, 355-56, 74 N.E. 603, 604, *appeal dismissed,* 199 U.S. 612 (1905). *See also* Rape v. Mobile & O. R.R., 136 Miss. 38, 100 So. 585 (1924), where the court, in finding a contract for "permanent" employment to be terminable at will, expressed a concern for unwary parties who might tie themselves up in perpetual contracts.

power of an employee's threat to quit his job. The situation of the employer differs drastically from that of the employee.[108] There is nothing more than the appeal of symmetry and a harkening back to hollow notions of mutuality to uphold any suggestion that the rights of employers must correspond to the rights of employees.

The cases involving interference with an employment relationship by a third party also provide a suitable basis for measuring damages in cases of abusive discharge. Generally speaking, the most significant item of damages in such cases has been the amount which the employee would have earned but for the defendant's wrongful act.[109] While it may not be sound to presume that the employment would have endured for the remainder of the employee's working life, the courts have deemed such a presumption fair because the defendant himself has made impossible any meaningful inquiry into what would have happened but for his wrongful act.

In cases of wrongful procurement the plaintiff is generally held not entitled to any damages which may have been or might yet be avoided by reasonably diligent efforts to obtain other employment.[110] Application of this principle in cases where the employee has been abusively discharged would be especially fitting, since the employee who is likely to suffer the least damage is also the one least affected by employer coercion. In other words, the employee who has enough mobility to avoid the consequences of his discharge will also have enough mobility to make him an unlikely target for oppression by the employer. But where the employee's experience is of special value only in his present employment or where his advanced age makes it doubtful that he can readily obtain comparable employment, he is more susceptible to improper exertion of the employer's power and less likely to succeed in mitigating damages.

There is, however, some authority to the effect that the employee is under no duty to avoid the consequences of his discharge by seeking other

---

108. As a matter of practical common sense, the situations of the employer . . . and that of one of its servants are very different. The loss or damage to the [employer] occasioned by the departure of one of its servants would, save in very exceptional circumstances, be negligible. To a servant . . . the security of employment . . . is of immense value.

McClelland v. N. I. Gen. Health Services Bd., [1957] 1 W.L.R. 594, 612.

109. *See, e.g.*, Hill Grocery Co. v. Carroll, 223 Ala. 376, 136 So. 789 (1931); Sullivan v. Barrows, 303 Mass. 197, 21 N.E.2d 275 (1939).

110. *See, e.g.*, Smetherham v. Laundry Workers' Local 75, 44 Cal. App. 2d 131, 111 P.2d 948 (1941); O'Brien v. Papas, 49 N.Y.S.2d 521 (Sup. Ct. 1944). Cases involving breach by the employer of an employment contract for a definite term are also instructive on the subject of the employee's duty to mitigate. *See, e.g.*, Beggs v. Dougherty Overseas, Inc., 287 F.2d 80 (2d Cir. 1961); Reinardy v. Bruzzese, 368 Mich. 688, 118 N.W.2d 952 (1962); Wood v. Ravenscroft, 135 Iowa 346, 112 N.W. 640 (1907). It has been held that by refusing to accept the employer's good faith offer of re-employment the employee fails to fulfill his duty to avoid the consequences of the employer's breach. Stevens v. Chicago Feather Co., 178 Ill. App. 455 (1913); Ryan v. Mineral County High School Dist., 27 Colo. App. 63, 146 P. 792 (1915); Schisler v. Perfection Milker Co., 193 Minn. 160, 258 N.W. 17 (1934). The employee need not, however, accept the offer of re-employment if anything degrading or offensive has occurred between the parties. *See* Hussey v. Holloway, 217 Mass. 100, 104 N.E. 471 (1914).

employment where the defendant has wrongfully procured his discharge.[111] This approach is punitive. If punishment is the objective of the remedy, explicitly punitive damages should be awarded.

The assessment of punitive damages[112] in cases of abusive firings would be fitting and desirable. They are typically awarded where the plaintiff's loss is caused by the defendant's "malicious, oppressive, willful, wanton, or reckless" behavior,[113] and the usual objective is deterrence.[114] Deterrence should also be a prime objective of the remedy for abusive discharge. The possibility of furthering this objective by the assessment of punitive damages provides an additional argument for a tort rather than a contract remedy. While punitive damages have been awarded in a number of cases involving tortious interference with the employment relationship by a third party,[115] they have generally not been recoverable against an employer in an action for breach of the employment contract, regardless of how evil the employer's motive in breaching.[116]

VIII. The Problem of Proof

As in any case which turns on motive or subjective intent, a cause of action for abusive discharge would pose difficult factual questions. In this difficulty lies perhaps the most cogent argument against giving every employee recourse to the courts in cases of abusive dismissal: that the danger of fictitious claims threatens interference with the normal exercise of the employer's right of discharge.

A libel action, *Mims v. Metropolitan Life Ins. Co.*,[117] throws light upon some of the significant aspects of this evidentiary problem. Prior to his dis-

111. Carmen v. Fox Film Corp., 204 App. Div. 776, 198 N.Y.S. 706 (1923).

112. *See generally* Morris, *Punitive Damages in Tort Cases*, 44 Harv. L. Rev. 1173 (1931). According to C. McCormick, Handbook on the Law of Damages 278-79 (1935), the allowance of punitive damages in tort cases is fully recognized in forty states and in the federal courts and is definitely rejected only in Louisiana, Massachusetts, Nebraska, and Washington.

113. Sebastian v. Wood, 246 Iowa 94, 101, 66 N.W.2d 841, 845 (1954).

114. "The award of such damages constitutes an effective deterrent to such offenders, and a salutary protection to society and the public in general." 246 Iowa at 100, 66 N.W.2d at 844.

115. *E.g.*, Harris v. Traders' & General Ins. Co., 82 S.W.2d 750 (Tex. Civ. App. 1935); Hilton v. Sheridan Coal Co., 132 Kan. 525, 297 P. 413 (1931); Hill Grocery Co. v. Carroll, 223 Ala. 376, 136 So. 789 (1931); Wyeman v. Deady, 79 Conn. 414, 65 A. 129 (1906). *Cf.* United States Fidelity & Guar. Co. v. Millonas, 206 Ala. 147, 89 So. 732 (1921), where the employee was allowed to recover damages against a third party for mental anguish caused by procurement of his discharge. Similarly, some cases allow "punitive damages," not as punishment, but as compensation for injured feelings. *See* Bixby v. Dunlap, 56 N.H. 456 (1816); Wise v. Daniel, 221 Mich. 229, 190 N.W. 746 (1922).

116. "Punitive damages are not recoverable for breach of contract." Restatement of Contracts § 342 (1932). Nor are damages for mental suffering ordinarily recoverable in a contract action. *See* Restatement of Contracts § 341 (1932); Holland v. Spartanburg Herald-Journal Co., 166 S.C. 454, 165 S.E. 203 (1932); Hazen v. Cobb, 96 Fla. 151, 117 So. 853 (1928); Addis v. Gramophone Co., [1909] A.C. 488.

117. 200 F.2d 800 (5th Cir. 1952), *cert. denied*, 345 U.S. 940 (1953).

missal the plaintiff had served the defendant insurance company for 32 years, the last 17 as branch manager. At issue in the suit was the defendant's explanation that the discharge was due to inefficiency and unsatisfactory production in the branches managed by the plaintiff. The plaintiff claimed that he had been discharged for his refusal to accede to a request, made in a chain letter sent by the defendant's supervisor of agencies, that $1.00 be contributed to a political campaign fund.

Similarly conflicting stories can be found in virtually every reported decision where the propriety of an employee's discharge is at issue.[118] The proof on both sides is likely to be equally biased, with testimony from a disinterested witness unavailable because the facts are contained within the employment relationship. So while the employee may testify to the employer's improper motives, the employer in most cases can counter with equally credible evidence of inefficiency, neglect or insubordination.

Ordinarily, where both sides present equally credible versions of the facts, the plaintiff will have failed to carry his burden. However, there is the danger that the average jury will identify with, and therefore believe, the employee. This possibility could give rise to vexatious lawsuits by disgruntled employees fabricating plausible tales of employer coercion. If the potential for vexatious suits by discharged employees is too great, employers will be inhibited in exercising their best judgment as to which employees should or should not be retained. And while as a matter of constitutional law the right of discharge is no longer absolute and inviolable, the employer's prerogative to make independent, good faith judgments about employees is important in our free enterprise system.

Compromise of the employer's power to make such judgments about professional, managerial or other high-ranking employees like the branch manager in the *Mims* case is especially undesirable. The higher ranking the employee, the more important to the success of the business is his effective performance. Compounding the potential for undue inhibition of the employer's judgment at the higher echelons of employment is the greater difficulty of articulating the basis for a discharge at that level. Compared to the wage earner, whose routine duties can generally be measured against a mechanical standard, the value of the salaried employee is more likely to be measured in such intangible qualities as imagination, initiative, drive, and personality. The employer's evaluation of the higher ranking employee is usually a highly personalized, intuitive judgment, and, as such, is more difficult to translate into concrete reasons which someone else—a juryman—can readily understand and appreciate. Indeed, even if it is conceded that the protection from unwarranted

118. *See, e.g.*, Seifert v. Arnold Bros., Inc., 138 Cal. App. 324, 31 P.2d 1059 (1934); Loos v. George Walter Brewing Co., 145 Wis. 1, 129 N.W. 645 (1911); United States Fidelity & Guar. Co. v. Millonas, 206 Ala. 147, 89 So. 732 (1921); Bell v. Faulkner, 75 S.W.2d 612 (Mo. App. 1934).

discharges afforded rank and file employees by labor agreements is appropriate, it might still be argued that no intrusion of any kind upon the employer's subjective evaluation of higher echelon employees should be tolerated. But this argument, when viewed against the strong interest in protecting the freedom and integrity of all employees, has force only if the sanctity of the normal right of discharge would be seriously impaired by unfounded claims of employer coercion.

The problem of proof is not insurmountable, for there are a number of evidentiary techniques available to the courts by which the genuineness of a claim might be reasonably guaranteed and serious infringement of the employer's normal right of discharge avoided.

Certainly, the employee should not be allowed to shift to the employer the burden of showing that the discharge was motivated by good cause by proving only that he capably performed the duties required by his job and was discharged for no apparent reason.[119] Such a standard would provide the employee with protection not only for his individual freedom but also for his interest in job security; such protection would restrict too greatly the employer's prerogative to dismiss an employee. The employee should be required in all cases to prove by affirmative and substantial evidence[120] that his discharge was actuated by reasons violative of his personal freedom or integrity.

The employee might even be held to a higher burden of proof than that normally required in civil actions. In that the proof offered in such cases would often be the word of one litigant against that of the other, the employee's action would be akin to others where the facts are likely to be known only to the litigants themselves.[121] In such lawsuits it has been customary to counteract the danger of deception by requiring proof of the issues by "clear and convincing" evidence.[122] Similarly, compelling the employee to bear this

---

119. Compare the following statement, made in connection with a complaint that an employee was discharged due to union activity in violation of the National Labor Relations Act:

> The Company does not have to prove nondiscrimination because of union activities. The [National Labor Relations] Board must prove discrimination because thereof. This burden on the Board to prove discrimination and to prove that discrimination was employed in the hiring or firing of a man because of his union activities does not shift from the Board.

Interlake Iron Corp. v. NLRB, 131 F.2d 129, 134 (7th Cir. 1942).

120. Compare the burden on the National Labor Relations Board "to prove affirmatively and by substantial evidence" that an employee was discharged because of his labor union activities. *See, e.g.*, NLRB v. Standard Coil Prods. Co., 224 F.2d 465, 470 (1st Cir. 1955); NLRB v. Reynolds Int'l Pen Co., 162 F.2d 680, 690 (7th Cir. 1947).

121. *E.g.*, Holley Coal Co. v. Globe Indem. Co., 186 F.2d 291 (4th Cir. 1950) (fraud); Steketee v. Steketee, 317 Mich. 100, 26 N.W.2d 724 (1947) (specific performance of oral contract); In re Mazanec's Estate, 204 Minn. 406, 283 N.W. 745 (1939) (undue influence); Jewell v. Allen, 188 Okla. 374, 109 P.2d 235 (1941) (action for damages for deceit). *See generally* 9 J. WIGMORE, EVIDENCE § 2498 (3d ed. 1940).

122. The litigant upon whom this burden of persuasion is placed should lose if the trier or triers of the fact are not convinced upon all the evidence that the facts upon which his claim depends . . . are highly probably true.

McBaine, *Burden of Proof: Degrees of Belief*, 32 CALIF. L. REV. 242, 254 (1944).

heavier burden of persuasion might yield appreciable protection from vexatious claims of abusive dismissal. Or, as has been done in other situations where the likelihood of spurious claims is great,[123] a claim of abusive dismissal might be rejected where there is nothing to corroborate it.

A somewhat similar barrier to unfounded claims might be established by buttressing the employer's inevitable counterargument—that the discharge was for good cause—by a presumption to that effect. Such a presumption would be particularly appropriate in the case of an employee who had not yet served a reasonably long time with the employer. Moreover, the burden of overcoming such a presumption would not have a harsh impact on the employee of short tenure, since he is unlikely to have as much at stake as the employee long in the service of one employer.

On the other hand, in cases like *Mims*, where the discharged employee had served the employer for 17 years as a branch manager and 32 years in all, a jury would probably be quite justified in finding little merit in an explanation that the plaintiff was fired for "chronic" inefficiency and incompetence. Indeed, there should be a point where long years of service cause the presumption of a properly motivated discharge to lose its force. The trier of fact might even properly consider each year's service beyond a reasonable probationary period as lending increasing strength to an inference that the employee's services were satisfactory to the employer. This would, to be sure, make it more difficult to discharge a long-tenured employee without substantial risk of a claim for abusive discharge, but a presumption of proper motives seems unwarranted in such cases. But there is something to be said for being more indulgent with the employee who invested many years with a particular employer and whose advanced age will make it difficult for him to obtain suitable employment elsewhere.[124] And as for the older employee who may be inclined to "coast" his last 10 or 15 years of work into retirement,[125] there will still be available to the employer a wide variety of sanctions, including discharge if all else fails, by which the employee's efficiency can be maintained.

---

123. Compare the rule allowing recovery in tort for mental suffering only when manifested by some visible physical symptom. Pennick v. Mirro, 189 F. Supp, 947 (E.D. Va. 1960); Espinosa v. Beverly Hosp., 114 Cal. App. 2d 232, 249 P.2d 843 (1952); Colla v. Mandella, 1 Wis. 2d 594, 85 N.W.2d 345 (1957). Compare also the cases where verdicts were overturned when based only on subjective testimony as to physical pain. Sporgis v. Butler, 40 Cal. App. 647, 181 P. 246 (1919); Johnson v. Great Northern Ry., 107 Minn. 285, 119 N.W. 1061 (1909).

124. It is well known that such indulgence is common practice among arbitrators deciding discharge cases under collective agreements. Very often an arbitrator will set aside the discharge of an employee with long service and freely admit that the same reasons would have justified the discharge of an employee with less seniority.

It has been pointed out that older workers tend to be less optimistic about their future prospects and tend to attach greater relative importance to job security than do younger workers. G. CROOK & M. HEINSTEIN, THE OLDER WORKER IN INDUSTRY 89-90 (1958). If this is true, the older employee is more susceptible to employer coercion—possibly a further reason for being more liberal in regard to the evidentiary sufficiency of an older employee's claim.

125. *Cf.* G. CROOK & M. HEINSTEIN, *supra* note 124, at 68, where the authors offer evidence that, contrary to the belief of some, job apathy does not set in with old age.

From a pragmatic standpoint, however, it is doubtful whether it would be necessary, or even desirable, for the courts to increase the burden of proof to protect the employer's normal right of discharge from undue intrusion. If the uncertainty of determining the employer's subjective intent should lead to findings generally unsatisfactory to employers, they would be quick to resort to arbitration or to create some other machinery for airing and adjusting such claims outside the courts. Thus a lack of confidence in courts and juries could lead, albeit in a roundabout way, to the creation of private means of settlement that might well be the most effective and expeditious way of handling such cases.[126]

IX. The Processes by Which the Limitation Might Be Established and Given Content

Some writers have boldly suggested that all the constitutional limitations imposed on the exercise of government power might be applied wholesale as checks on private organizations, chiefly corporations, whose power is quasi-governmental.[127] This proposal draws support not only from the equation of governmental and corporate power, but also from an increasing identification of industrial and governmental interests stemming, in part, from the large number of government projects contracted out to private industry.[128] But while there may be some collaboration and some similarity in size between government and private corporations, any suggestion that the Bill of Rights and the fourteenth amendment as such will be applied to large corporations seems visionary.[129] And while the theory usually offered in support of the

126. Compare the strong preference which employers and unions have for arbitrating, rather than litigating, discharge cases and other grievances arising under a collective bargaining agreement. *See* Jones & Smith, *Management and Labor Appraisals and Criticisms of the Arbitration Process: A Report with Comments*, 62 Mich. L. Rev. 1115 (1964).

127. The historical trend of judicial decision-making has been to bring more and more activity within the reach of the limitations of the Constitution. Since 1789, moreover, more and more governmental activity has been made subject to due process and similar limitations. The next logical step would be to draw private governments into the tent of state action.

A. S. Miller, Private Governments and the Constitution 13 (Occasional Paper for Center for the Study of Democratic Institutions, 1959).

128. *See* J. K. Galbraith, The New Industrial State 296-317 (1967).

129. There has been strong criticism of the suggestion that sectors of private power might be regulated by the Constitution. *See, e.g.*, Wellington, *The Constitution, the Labor Union, and "Governmental Action,"* 70 Yale L.J. 345, 348 (1961):

> Undue fascination with the supposed structural and power similarities unions and corporations have with government can be misleading. Some commentators have become so fascinated, and it is this fascination that provides the decisive reason for their advocacy of regulation by the Constitution. There is also little doubt that some commentators are motivated by an undisciplined desire to "let the mind be bold." This sort of thinking deserves little sympathy. The Bill of Rights and the fourteenth amendment are the great instruments with which courts protect the people from misused governmental power. The view that because unions and corporations are somehow similar to government they too should be restrained by these same constitutional provisions has perhaps an aesthetic and emotional appeal. Its analytical shortcomings, however, are fatal.

See also Wechsler, *Toward Neutral Principles of Constitutional Law*, 73 Harv. L. Rev. 1, 31 (1959).

proposal—that corporations operate under state charters and are therefore state agencies[130]—has a superficial appeal, untold difficulties would inevitably be encountered in attempting to fit widely differing private relationships onto a Procrustean bed of constitutional principles. The corporate charter theory, moreover, leaves free from limitations a large number of unincorporated firms and establishments with great economic power. The suggestion has apparently been made on the assumption that the legislative and common law processes are too slow moving to meet the increasing need to protect individual freedom from oppression by private entities.[131] If contrary to this assumption, the legislatures or the courts proceed quickly with the task of developing other approaches, it will not be necessary to resort to the drastic yet inadequate step of limiting the exercise of private power through recourse to constitutional law.[132]

The states could, of course, enact more statutes prohibiting an employer from coercing an employee in specific ways, such as interfering with his political free choice, compromising his professional ethics, and coercing him to engage in fraud or criminal activity. Most important, such legislation could give the discharged employee a remedy in damages for his resulting injuries. But because it would be impossible to anticipate all the particular aspects of the employee's life which might be subjected to abuse, and since what is reasonable for an employer to demand of an employee might vary from job to job, a single, general statutory definition of the wrong might be preferable. If statutes could be passed which identify the wrong as, for example, interference with the freedom or integrity of the employee in respects which bear no reasonable relationship to the employment or which do not advance the legitimate interests of the employer,[133] the courts could be left to perform the

130. One may reasonably forecast, in the future, direct application of constitutional limitations to the corporation, merely because it holds a state charter and exercises a degree of economic power sufficient to make its practices "public" rules.

Berle, *Constitutional Limitations on Corporate Activity—Protection of Personal Rights from Invasion Through Economic Power*, 100 U. PA. L. REV. 933, 953 (1952).

[W]e could consider applying to the corporation *the whole pattern of controls* laid upon the states when the Federal Republic was created under the Constitution of 1787. The parallel is not too fanciful since many of the states in fact started out as corporations created by the Crown, bodies politic endowed with public authority.

Latham, *The Commonwealth of the Corporation*, 55 NW. U.L. REV. 25, 35 (1960) (emphasis in original).

131. Another slower and less predictable line of protection is to rely on the growth of the common law, particularly the law of torts.

Berle, *Legal Problems of Economic Power*, 60 COLUM. L. REV. 4, 9 (1960).

132. *Cf.* Wellington, *The Constitution, the Labor Union, and "Governmental Action,"* 70 YALE L.J. 345, 348 (1961):

The need to regulate unions and corporations is undeniable; but it need not be assumed a priori that the Constitution is the proper regulatory instrument. Other, more appropriate, means may be available to accomplish the same desirable ends.

133. Compare the standard applied in Mitchell v. International Ass'n of Machinists, 196 Cal. App. 2d 796, 16 Cal. Rptr. 813 (1961), where two members were expelled from a labor union for actively campaigning for the enactment of a state right-to-work law.

function, for which they are well suited, of giving reasoned elaboration to a broad statutory provision.[134] And through a process of continuing judicial articulation a detailed bill of rights for all employees could be provided.

Instead of having claims of abusive dismissal heard by courts and juries, legislation could vest exclusive authority to decide such cases in the various civil rights or fair employment practices commissions which have been established in a number of jurisdictions.[135] So expanding the authority of these commissions would seem to be a most appropriate means of dealing with the general problem of employer interference and oppression.[136] This approach would provide the advantages of administrative responsibility and expertise in a limited area, for the task of such a commission in determining the employer's motive would be essentially the same in cases of abusive discharge as in cases where discrimination is claimed. Also, the range of remedies available to such commissions is typically wider than that at the disposal of the courts,[137] a difference which might allow the commission to deal with abusive employer practices in contexts other than discharge. Such an expansion would obviously be in keeping with the commissions' function of protecting the "civil rights" of individuals from abuse at the hands of private employers. Indeed, it seems anomalous that these tribunals provide relief to an employee who is discharged because of his race or religion yet do not grant similar relief to an employee who is discharged because he exercised his right of free speech or because he refused to commit some fraud or crime at his employer's behest.

The problem does seem to be one suited to legislative inquiry and solution.[138] As a practical matter, however, the prospects for any kind of general

---

The court ordered their reinstatement, saying that "where the political activity of the member is not patently in conflict with the union's best interests, the union should not be permitted to use its power over the individual to curb the advocacy of his political views." *Id.* at 807, 16 Cal. Rptr. at 819-20.

134. As Chief Justice Hughes once said of the Sherman Antitrust Act, such a standard of reasonableness, "[a]s a charter of freedom," would have a "generality and adaptability comparable to that found to be desirable in constitutional provisions." Appalachian Coals, Inc. v. United States, 288 U.S. 344, 359-60 (1933).

135. The most pervasive enactment in this area, of course, is the Federal Civil Rights Act of 1964, Title VII of which establishes an Equal Employment Opportunity Commission vested with authority to prevent discrimination in employment on the basis of race, color, religion, sex or national origin. 42 U.S.C. § 2000e et seq. (1964). The Act covers almost all employers of 25 or more employees as well as employment agencies and labor unions. *See generally* Berg, *Equal Employment Opportunity Under the Civil Rights Act of 1964*, 31 BKLYN. L. REV. 62 (1964). Thirty-one states now have administrative tribunals which enforce prohibitions against discrimination in employment. *See* Bonfield, *The Origin and Development of American Fair Employment Legislation*, 52 IOWA L. REV. 1043, 1088 n.208 (1967).

136. *Cf.* Note, *The Right to Equal Treatment: Administrative Enforcement of Anti-discrimination Legislation*, 74 HARV. L. REV. 526 (1961).

137. Under Title VII of the Federal Civil Rights Act of 1964, for example, the Equal Employment Opportunity Commission can grant injunctive relief and such other relief "as may be appropriate," which may include reinstatement and/or back pay. 42 U.S.C. § 2000e-5(g) (1964). Anglo-American courts, on the other hand, have traditionally refused to reinstate a wrongfully dismissed employee, or, as they put it, to order specific performance of a personal service contract. Ogden v. Fossick, 32 L.J.H.L. (n.s.) 73 (1863); McMenamin v. Philadelphia Trans. Co., 356 Pa. 88, 51 A.2d 702 (1947).

138. *Cf.* R. HALE, FREEDOM THROUGH LAW 548 (1952):

legislative reform in this area are dim. The obstacles which commonly hinder legislative reforms of this sort have been commented upon elsewhere.[139] Suffice it to say that general statutory limitations on the employer's right of discharge are unlikely to be enacted so long as there is no strong lobby to promote them.[140] Employees having diverse job specialties and working at varying echelons of employment simply are not equipped to form a cohesive group with enough power to influence legislators.[141] The unlikelihood that such legislation will be enacted in the foreseeable future is enhanced by the strong interest groups to be counted on to oppose it.[142] One need not be an extreme cynic to say that employers would not favor such legislation. Nor could organized labor be expected to favor laws which would give individual employees a means of protecting themselves without need of a union. Therefore, it appears that protection of all employees from the abusive exercise of employer power will have to originate, if it is to be established at all, in the courts.[143]

> As far as individual liberty is concerned, it is just as important that legislative bodies should be able to protect persons from oppression at the hands of private groups which exercise power indistinguishable from that exercised by government as it is that courts should be able to protect them from oppression by officials whose power is more generally recognized as governmental.

139. *See, e.g.*, Peck, *The Role of the Courts and Legislatures in the Reform of Tort Law*, 48 MINN. L. REV. 265 (1963). Among the characteristics of the legislative process which the author points to as generally obstructing statutory reform of tort law are that legislators are indifferent, lack insight and experience, are paid inadequate wages, fail to hold satisfactory committee and public hearings, and are subject to well-organized lobbies and pressure groups.

140. Professional associations might be the most logical organizations to seek such legislation, but at least one observer has noted that these groups have so far avoided undertaking to protect their constituents from arbitrary employer power. Barkin, *supra* note 33, at 240.

141. Compare the strong lobby of the automobile dealers that led to the enactment of the Federal Automobile Dealer Franchise Act of 1956. *Cf. supra* note 55. Consider, in this connection, the more broadly based community of interests among automobile dealers than among employees as a class.

142. Consider, in this connection, the theory that the influential legislative lobbyists of today represent "a series of groups, each of which has struggled for and finally attained *a power to stop things conceivably inimical to its interests* and, within far narrower limits, to start things." D. RIESMAN, THE LONELY CROWD 244 (1950) (emphasis added).

See also D. TRUMAN, THE GOVERNMENTAL PROCESS 362-63 (1951); Dykstra, *The Impact of Pressure Groups on the Legislative Process*, 1951 WASH. U.L.Q. 306.

143. *Cf.* Keeton, *Judicial Law Reform—A Perspective on the Performance of Appellate Courts*, 44 TEXAS L. REV. 1254, 1260 (1966):

> They assert . . . that it is the province of the legislature, not the court, to meet the need for reform. . . . [T]he assertion that legislatures could perform this task, if only they would, grows ever more questionable in our changing social and political order. Even if legislatures could do the task alone, moreover, their practice has conclusively demonstrated that they do not. If courts also fail to act when law reform is needed, the consequence will be a constantly increasing heritage of outmoded rules.

Because of the obstacles to legislative reform of private law, particularly in tort law (see note 139 *supra*), many commentators have urged a more active and creative role for the judiciary. *See, e.g.*, W. SEAVEY, COGITATIONS ON TORTS 52-72 (1954); Keeton, *Creative Continuity in the Law of Torts*, 75 HARV. L. REV. 463 (1962); James, *Tort Law in Midstream: Its Challenge to the Judicial Process*, 8 BUFFALO L. REV. 315 (1959); Peck, *The Role of the Courts and Legislatures in the Reform of Tort Law*, 48 MINN. L. REV. 265 (1963).

For a survey of various philosophical and jurisprudential attitudes toward judicial creativity, *see* Friedmann, *Legal Philosophy and Judicial Lawmaking*, 61 COLUM. L. REV. 821 (1961).

CONCLUSION

The judiciary has not been reluctant to expand the meaning of constitutional provisions in order to protect the individual from governmental oppression. It is something of a paradox that the courts have so far displayed no similar bent for invention and improvisation when it comes to protecting individuals, particularly in their highly vulnerable status as employees, from the private establishments upon which they are becoming increasingly dependent. Instead, there has been a blind acceptance of the employer's absolute right of discharge. This outmoded doctrine has been supported by technical principles of contract law. But another common law resource can be tapped to eliminate much of the potential for oppression inherent in the employment relationship. Through adaptation of the general emphasis on wrongful or ulterior motives which today pervades the law of torts the courts could fashion a remedy for the abusively discharged employee and thereby give to *all* employees some assurance that they will be their own masters as to matters not their employers' business.

This page intentionally left blank

In Defense of the Contract at Will

Richard A. Epstein†

The persistent tension between private ordering and government regulation exists in virtually every area known to the law, and in none has that tension been more pronounced than in the law of employer and employee relations. During the last fifty years, the balance of power has shifted heavily in favor of direct public regulation, which has been thought strictly necessary to redress the perceived imbalance between the individual and the firm. In particular the employment relationship has been the subject of at least two major statutory revolutions. The first, which culminated in the passage of the National Labor Relations Act in 1935,[1] set the basic structure for collective bargaining that persists to the current time. The second, which is embodied in Title VII of the Civil Rights Act of 1964,[2] offers extensive protection to all individuals against discrimination on the basis of race, sex, religion, or national origin. The effect of these two statutes is so pervasive that it is easy to forget that, even after their passage, large portions of the employment relation remain subject to the traditional common law rules, which when all was said and done set their face in support of freedom of contract and the system of voluntary exchange. One manifestation of that position was the prominent place that the common law, especially as it developed in the nineteenth century, gave to the contract at will. The basic position was well set out in an oft-quoted passage from *Payne v. Western & Atlantic Railroad:*

> [M]en must be left, without interference to buy and sell where they please, and to discharge or retain employees at will for good cause or for no cause, or even for bad cause without thereby being guilty of an unlawful act *per se.* It is a right which an employe may exercise in the same way, to the same

---

† James Parker Hall Professor of Law, University of Chicago.

[1] Act of July 5, 1935, ch. 372, 49 Stat. 449 (codified as amended at 29 U.S.C. §§ 151-169 (1982)).

[2] Pub. L. No. 88-352, 78 Stat. 253 (codified as amended at 42 U.S.C. §§ 2000e to 2000e-17 (1982)).

extent, for the same cause or want of cause as the employer.[3]

The survival of the contract at will, and the frequency of its use in private markets, might well be taken as a sign of its suitability for employment relations. But the contract at will has been in retreat even at common law, as the movement for public control of labor markets has now spilled over into the judicial arena. The judicial erosion of the older position has been spurred on by academic commentators, who have been almost unanimous in their condemnation of the at-will relationship, often treating it as an archaic relic that should be jettisoned along with other vestiges of nineteenth-century laissez-faire.[4] Thus it is commonly asserted

---

[3] Payne v. Western & Atl. R.R., 81 Tenn. 507, 518-19 (1884), *overruled on other grounds*, Hutton v. Watters, 132 Tenn. 527, 544, 179 S.W. 134, 138 (1915). The passage continues as follows:

> He may refuse to work for a man or company, that trades with any obnoxious person, or does other things which he dislikes. He may persuade his fellows and the employer may lose all his hands and be compelled to close his doors; or he may yield to the demand and withdraw his custom or cease his dealings, and the obnoxious person be thus injured or wrecked in business.

81 Tenn. at 519. It should be noted that *Payne* did not itself involve the discharge of an employee for a bad reason or no reason at all. As the last two quoted sentences indicate, the question of the status of the contract arose obliquely, in a defamation suit by a merchant against a railroad. The railroad's yard master had posted a sign that read: "Any employe of this company on Chattanooga pay-roll who trades with L. Payne from this date will be discharged. Notify all in your department." *Payne*, 81 Tenn. at 510.

The plaintiff Payne claimed that his business, which had been heavily dependent upon the trade of railroad workers, had thereby been ruined. The court held for the defendant on the grounds that (a) there was no defamation implicit in the announcement and (b) the employer's notice to its employees was within its rights because all the contracts with its workers were terminable at will. *Hutton* overruled *Payne*, not on the ground that contracts at will were against public policy, but on an abuse-of-rights theory according to which an employer cannot use his right to discharge employees for the sole purpose of harming third-party interests. The propriety of the *Hutton* theory is a difficult question, but my views tend toward those of the *Payne* court. *See* Epstein, *A Common Law for Labor Relations: A Critique of the New Deal Labor Legislation*, 92 YALE L.J. 1357, 1367-69, 1381 (1983).

[4] *E.g.*, Blackburn, *Restricted Employer Discharge Rights: A Changing Concept of Employment at Will*, 17 AM. BUS. L.J. 467, 491-92 (1980); Blades, *Employment at Will v. Individual Freedom: On Limiting the Abusive Exercise of Employer Power*, 67 COLUM. L. REV. 1404, 1405-06, 1413-14, 1435 (1967); Blumrosen, *Employer Discipline: United States Report*, 18 RUTGERS L. REV. 428, 428-34 (1964); Feinman, *The Development of the Employment at Will Rule*, 20 AM. J. LEGAL HIST. 118, 131-35 (1976); Murg & Scharman, *Employment at Will: Do the Exceptions Overwhelm the Rule?*, 23 B.C.L. REV. 329, 338-40, 383-84 (1982); Peck, *Unjust Discharges from Employment: A Necessary Change in the Law*, 40 OHIO ST. L.J. 1, 1-10 (1979); Summers, *Individual Protection Against Unjust Dismissal: Time for a Statute*, 62 VA. L. REV. 481, 484 (1976); Weynard, *Present Status of Individual Employee Rights*, PROC. N.Y.U. 22D ANN. CONF. ON LAB. 171, 214-16 (1970); Note, *Guidelines for a Public Policy Exception to the Employment at Will Rule*, 13 CONN. L. REV. 617, 641-42 (1980); Note, *Protecting Employees at Will Against Wrongful Discharge: The Public Policy Exception*, 96 HARV. L. REV. 1931, 1931-35 (1983); Note, *Protecting At Will Employees Against Wrongful Discharge: The Duty to Terminate Only in Good Faith*, 93

that, however congenial the contract at will might have been to the conditions of earlier times, major transformations in firm organization and industrial production have rendered it anachronistic today. One early and influential attack on the contract at will shows the importance of the issues that it raises:

> It is a widely accepted proposition that large corporations now pose a threat to individual freedom comparable to that which would be posed if governmental power were unchecked. The proposition need not, however, be limited to the mammoth business corporation, for the freedom of the individual is threatened whenever he becomes dependent upon a private entity possessing greater power than himself. Foremost among the relationships of which this generality is true is that of employer and employee.[5]

The contract at will is thus thought to be particularly unwise because it invites the exercise of arbitrary power by persons with a dominant economic position against individuals whose mobility is said to be limited by the structure of labor markets. The absence of viable alternative employment opportunities is thought to leave employees vulnerable to coercion and exploitation. In the extreme situation, employers (or their managers) are said to fire workers out of personal animosity; the animosity may stem from the worker's refusal to grant personal or sexual favors or from a simple and irrational dislike of the personal characteristics or habits of the employee.[6] Once the outcomes of these market imperfections are identified, the conclusion is said to follow swiftly: where discharges from employment are not made for sound economic reasons, or to advance the financial interest of the firm, the power of the law must be used to redress the perceived imbalance.

The courts have been heavily influenced by these arguments,

HARV. L. REV. 1816, 1824-28 (1980) [hereinafter cited as Note, *Wrongful Discharge*]; Note, *A Common Law Action for the Abusively Discharged Employee*, 26 HASTINGS L.J. 1435, 1443-46 (1975); Note, *Implied Contract Rights to Job Security*, 26 STAN. L. REV. 335, 337-40 (1974); Note, *California's Controls on Employer Abuse of Employee Political Rights*, 22 STAN. L. REV. 1015, 1015-20 (1970).

[5] Blades, *supra* note 4, at 1404 (footnotes omitted). Examples of the subsequent literature on the point are cited *supra* note 4.

[6] Such allegations were made in Alcorn v. Anbro Eng'g, Inc., 2 Cal. 3d 493, 496-97, 468 P.2d 216, 217, 86 Cal. Rptr. 88, 89 (1970) (employee dismissed due to employer's animosity toward employee on the basis of his race and his union activities); Monge v. Beebe Rubber Co., 114 N.H. 130, 131-32, 316 A.2d 549, 550-51 (1974) (employee dismissed due to refusal to date her foreman); Hutton v. Watters, 132 Tenn. 527, 529-30, 179 S.W. 134, 134-35 (1915) (employees dismissed for patronizing a boarding house whose owner was personally disliked by employer).

and the tempo of their intervention into market processes has increased rapidly in recent years.[7] The underlying rationale for the intervention is well captured by the conclusion of the New Hampshire Supreme Court in *Monge v. Beebe Rubber Co.*: "We hold that a termination by the employer of a contract of employment at will which is motivated by bad faith or malice or based on retaliation is not in the best interest of the economic system or the public good and constitutes a breach of the employment contract."[8] Once the wrongful discharge is established, damages can be awarded or reinstatement can be ordered. More recently, legislatures have begun to consider proposals that replace the contract at will with an action for wrongful discharge, so that all private contracts of employment would be terminable by the employer only upon a showing of

---

[7] There appear to be three theories under which courts have recognized a cause of action for wrongful discharge: (1) discharge violates public policy; (2) discharge violates an implied term of contract; and (3) discharge constitutes a tort of emotional harm.

The public-policy approach is the most widely accepted. *See, e.g.*, Tameny v. Atlantic Richfield Co., 27 Cal. 3d 167, 176-77, 610 P.2d 1330, 1336, 164 Cal. Rptr. 839, 844-45 (1980) (dismissal for refusal to participate in price-fixing scheme); Glenn v. Clearman's Golden Cock Inn, Inc., 192 Cal. App. 2d 793, 796-97, 13 Cal. Rptr. 769, 771 (1961) (employee dismissed for joining a union); Petermann v. Teamsters Local 396, 174 Cal. App. 2d 184, 188-89, 344 P.2d 25, 27-28 (1959) (employee dismissed for refusing to commit perjury); Kelsay v. Motorola, Inc., 74 Ill. 2d 172, 181-83, 384 N.E.2d 353, 358-59 (1978) (employee dismissed for filing workers' compensation claim); Frampton v. Central Ind. Gas Co., 260 Ind. 249, 252-53, 297 N.E.2d 425, 428 (1973) (same); Sventko v. Kroger Co., 69 Mich. App. 644, 648, 245 N.W.2d 151, 153 (1976) (same); O'Sullivan v. Mallon, 160 N.J. Super. 416, 417-18, 390 A.2d 149, 150 (Law Div. 1978) (medical technician fired for refusal to perform operation for which she was not licensed); Nees v. Hocks, 272 Or. 210, 218-19, 536 P.2d 512, 515-16 (1975) (employee dismissed for serving on a jury); Reuther v. Fowler & Williams, Inc., 255 Pa. Super. 28, 32, 386 A.2d 119, 120 (1978) (dismissal for serving on jury) (dictum); Harless v. First Nat'l Bank, 246 S.E.2d 270, 275-76 (W. Va. 1978) (employee dismissed for calling attention to employer's violation of law).

The contract theory has also been used by a number of courts. *See, e.g.*, McKinney v. National Dairy Council, 491 F. Supp. 1108, 1122 (D. Mass. 1980) (dismissal due to age); Cleary v. American Airlines, 111 Cal. App. 3d 443, 456, 168 Cal. Rptr. 722, 729 (1980) (dismissal without the investigation or hearing that was normally part of employer's procedure); Maddaloni v. Western Mass. Bus Lines, 386 Mass. 877, 878-80, 438 N.E.2d 351, 354-56 (1982) (discharge to avoid paying commissions due); Fortune v. National Cash Register Co., 373 Mass. 96, 101, 364 N.E.2d 1251, 1255-56 (1977) (same); Toussaint v. Blue Cross & Blue Shield, 408 Mich. 579, 598, 292 N.W.2d 880, 885 (1980) (implied term forbidding dismissal except for cause); Weiner v. McGraw-Hill, Inc., 57 N.Y.2d 458, 465-66, 443 N.E.2d 441, 445, 457 N.Y.S.2d 193, 197 (1982) (same).

Finally, courts have occasionally treated discharge as a tortious infliction of emotional harm. *See, e.g.*, Alcorn v. Anbro Eng'g, Inc., 2 Cal. 3d 493, 498, 468 P.2d 216, 218, 86 Cal. Rptr. 88, 90-91 (1970); Agis v. Howard Johnson Co., 371 Mass. 140, 144-45, 355 N.E.2d 315, 318-19 (1976).

[8] 114 N.H. 130, 133, 316 A.2d 549, 551 (1974). Note, however, that the New Hampshire Supreme Court has since retreated from this broad view. *See* Howard v. Dorr Woolen Co., 120 N.H. 295, 297, 414 A.2d 1273, 1274 (1980) (limiting the *Monge* approach to cases involving violations of specific public policies).

"just cause," which would often be determined by an elaborate structure of mediation, arbitration, and administrative action.[9]

There is thus today a widely held view that the contract at will has outlived its usefulness. But this view is mistaken. The contract at will is not ideal for every employment relation. No court or legislature should ever command its use. Nonetheless, there are two ways in which the contract at will should be respected: one deals with entitlements against regulation and the other with presumptions in the event of contractual silence.

First, the parties should be permitted as of right to adopt this form of contract if they so desire. The principle behind this conclusion is that freedom of contract tends both to advance individual autonomy and to promote the efficient operation of labor markets.

Second, the contract at will should be respected as a rule of construction in response to the perennial question of gaps in contract language: what term should be implied in the absence of explicit agreement on the question of duration or grounds for termination? The applicable standard asks two familiar questions: what rule tends to lend predictability to litigation and to advance the joint interests of the parties?[10] On both these points I hope to show that the contract at will represents in most contexts the efficient solution to the employment relation. To be sure, the stakes are lower where the outright prohibition is no longer in the offing. No rule of construction ever has the power of a rule of regulation, since the parties by negotiation can reverse what the law otherwise commands. Nonetheless, bad rules of contract construction have costs that should not be understated, here or elsewhere. The rule of construction is normally chosen because it reflects the dominant practice in a given class of cases and because that practice is itself regarded as making good sense for the standard transactions it governs. It is of course freely waivable by a joint expression of contrary intention. When the law introduces a just-cause requirement,

---

[9] Bills have been introduced in Michigan, California, Wisconsin, Ohio, Pennsylvania, and Massachusetts. Chicago Sun-Times, June 10, 1984, at 32, col. 1 (2-star ed.).

[10] The traditional rule has been codified under current California law: "An employment, having no specified term, may be terminated at the will of either party on notice to the other." CAL. LAB. CODE § 2922 (West 1971). Indeed, this should mean, as it now does, that where a contract speaks of "permanent" employment, the presumption should again be that the contract is terminable at will, for all that "permanent" connotes is the absence of any definite termination date. It does not imply one in which there is a lifetime engagement by either employer or employee, especially where none of the subsidiary terms for such a long-term relationship is identified by the parties. The proper rule of construction should be that the contract is terminable at will by either side.

it flies in the face of ordinary understandings and thus rests upon an assumption that just-cause arrangements are in the broad run of cases either more frequent or desirable than the contract at will, though neither is the case. Where this rule of construction is used, therefore, contracting-out will have to take place in the very large number of cases where the parties desire to conform to the norm by entering into a contract at will. Furthermore, it may be difficult to waive the for-cause requirement in fact, even if waiver is formally allowable as a matter of law, because of high standards for "informed" waiver that cannot be met after the fact. By degrees, the original presumption against the contract at will could so gain in strength that a requirement that is waivable in theory could easily become conclusive in fact.

These complications should all be avoided. The critics of the contract at will all point out imperfections in the current institutional arrangements, but they do not take into account the nonlegal means of preserving long-term employment relationships, and they ignore the greater imperfections that are created under alternative legal rules. Contracts at will are consistent with public policy and should be welcomed by it, not because they are perfect, but because in many contexts they respond to the manifold perils of employment contracts better than any rivals that courts or legislatures can devise.

In this area of private-contracting autonomy, there are some exceptions, arising out of the infrequent cases in which discharge of the contract at will is inconsistent with the performance of some public duty or with the protection of some public right. Just as a contract to commit murder should not be enforceable, neither should one to pollute illegally or to commit perjury.[11] But these

[11] This problem has arisen where employees at will have refused to perjure themselves on behalf of the employer, *e.g.*, Petermann v. Teamsters Local 396, 174 Cal. App. 2d 184, 188-89, 344 P.2d 25, 27-28 (1959) (discharge for refusal to commit perjury held wrongful), or where workers have been dismissed because they have filed workers' compensation claims, *e.g.*, Frampton v. Central Ind. Gas Co., 260 Ind. 249, 252-53, 297 N.E. 2d 425, 428 (1973) (discharge for filing claim held wrongful). It seems clear that any contract to commit perjury should simply be treated as illegal. The workers' compensation case is more difficult both because there is less justification for the coercive character of compensation, since no third-party interests are at stake, and because in all events the worker is entitled to file his claim and will do so if its value exceeds the gains he expects from the employment contract. A common law court cannot, however, attack the soundness of a statutory compensation system, so that this restraint on freedom of contract should be as valid as one imposed for the protection of strangers. At this point the central question concerns the proper remedy. Typically, reinstatement of the plaintiff is ordered, which has the disadvantage of requiring the court to supervise an ongoing relationship. It may well be that the employer should be able to fire the worker, but nonetheless be required to pay damages, preferably fixed by statute,

cases, however difficult in their own right, in no way require abandoning the basic common law presumption in favor of contracts at will. The recent efforts to undermine or abolish the contract at will should be evaluated not in terms of what they *hope* to achieve, whether stated in terms of worker participation, industrial harmony, fundamental fairness, or enlightened employment relations. Instead they should be evaluated for the generally harsh results that they actually produce. They introduce an enormous amount of undesirable complexity into the law of employment relations; they increase the frequency of civil litigation; and over the broad run of cases they work to the disadvantage of both the employers and the employees whose conduct they govern.

In the remainder of this paper, I examine the arguments that can be made for and against the contract at will. I hope to show that it is adopted not because it allows the employer to exploit the employee, but rather because over a very broad range of circumstances it works to the mutual benefit of both parties, where the benefits are measured, as ever, at the time of the contract's formation and not at the time of dispute. To justify this result, I examine the contract in light of the three dominant standards that have emerged as the test of the soundness of any legal doctrine: intrinsic fairness, effects upon utility or wealth, and distributional consequences. I conclude that the first two tests point strongly to the maintenance of the at-will rule, while the third, if it offers any guidance at all, points in the same direction.

I. The Fairness of the Contract at Will

The first way to argue for the contract at will is to insist upon the importance of freedom of contract as an end in itself. Freedom of contract is an aspect of individual liberty, every bit as much as freedom of speech, or freedom in the selection of marriage partners or in the adoption of religious beliefs or affiliations. Just as it is regarded as prima facie unjust to abridge these liberties, so too is it presumptively unjust to abridge the economic liberties of individuals. The desire to make one's own choices about employment may be as strong as it is with respect to marriage or participation in religious activities, and it is doubtless more pervasive than the desire to participate in political activity. Indeed for most people, their own health and comfort, and that of their families, depend

---

to the worker. On the clear public-policy exception to the validity of contracts at will, see generally Murg & Scharman, *supra* note 4, at 343-55.

critically upon their ability to earn a living by entering the employment market. If government regulation is inappropriate for personal, religious, or political activities, then what makes it intrinsically desirable for employment relations?

It is one thing to set aside the occasional transaction that reflects only the momentary aberrations of particular parties who are overwhelmed by major personal and social dislocations. It is quite another to announce that a rule to which vast numbers of individuals adhere is so fundamentally corrupt that it does not deserve the minimum respect of the law. With employment contracts we are not dealing with the widow who has sold her inheritance for a song to a man with a thin mustache. Instead we are dealing with the routine stuff of ordinary life; people who are competent enough to marry, vote, and pray are not unable to protect themselves in their day-to-day business transactions.

Courts and legislatures have intervened so often in private contractual relations that it may seem almost quixotic to insist that they bear a heavy burden of justification every time they wish to substitute their own judgment for that of the immediate parties to the transactions. Yet it is hardly likely that remote public bodies have better information about individual preferences than the parties who hold them. This basic principle of autonomy, moreover, is not limited to some areas of individual conduct and wholly inapplicable to others. It covers all these activities as a piece and admits no ad hoc exceptions, but only principled limitations.

This general proposition applies to the particular contract term in question. Any attack on the contract at will in the name of individual freedom is fundamentally misguided. As the Tennessee Supreme Court rightly stressed in *Payne*, the contract at will is sought by both persons.[12] Any limitation upon the freedom to enter into such contracts limits the power of workers as well as employers and must therefore be justified before it can be accepted. In this context the appeal is often to an image of employer coercion.[13] To be sure, freedom of contract is not an absolute in the employment context, any more than it is elsewhere. Thus the principle must be understood against a backdrop that prohibits the use of private contracts to trench upon third-party rights, including uses that interfere with some clear mandate of public policy, as

---

[12] Payne v. Western & Atl. R.R., 81 Tenn. 507, 518-19 (1884). *See supra* note 3 and accompanying text.

[13] *See supra* note 5 and accompanying text.

in cases of contracts to commit murder or perjury.[14]

In addition, the principle of freedom of contract also rules out the use of force or fraud in obtaining advantages during contractual negotiations; and it limits taking advantage of the young, the feeble-minded, and the insane.[15] But the recent wrongful discharge cases do not purport to deal with the delicate situations where contracts have been formed by improper means or where individual defects of capacity or will are involved. Fraud is not a frequent occurrence in employment contracts, especially where workers and employers engage in repeat transactions. Nor is there any reason to believe that such contracts are marred by misapprehensions, since employers and employees know the footing on which they have contracted: the phrase "at will" is two words long and has the convenient virtue of meaning just what it says, no more and no less.[16]

An employee who knows that he can quit at will understands what it means to be fired at will, even though he may not like it after the fact. So long as it is accepted that the employer is the full owner of his capital and the employee is the full owner of his labor, the two are free to exchange on whatever terms and conditions they see fit, within the limited constraints just noted. If the arrangement turns out to be disastrous to one side, that is his problem; and once cautioned, he probably will not make the same mistake a second time. More to the point, employers and employees are unlikely to make the same mistake once. It is hardly plausible that contracts at will could be so pervasive in all businesses and at all levels if they did not serve the interests of employees as well as employers. The argument from fairness then is very simple, but not for that reason unpersuasive.

II. The Utility of the Contract at Will

The strong fairness argument in favor of freedom of contract makes short work of the various for-cause and good-faith restrictions upon private contracts. Yet the argument is incomplete in several respects. In particular, it does not explain why the presumption in the case of silence should be in favor of the contract at

---

[14] *See supra* note 11.

[15] For my elaboration of this general point, see Epstein, *Unconscionability: A Critical Reappraisal*, 18 J.L. & Econ. 293 (1975).

[16] In the absence of force or fraud, any disclosure law would be regarded as only a nuisance by employers and employees alike, whatever the case for such laws in other contexts. *See, e.g.*, Kronman, *Mistake, Disclosure, Information, and the Law of Contracts*, 7 J. Legal Stud. 1, 11-18 (1978).

will. Nor does it give a descriptive account of *why* the contract at will is so commonly found in all trades and professions. Nor does the argument meet on their own terms the concerns voiced most frequently by the critics of the contract at will. Thus, the commonplace belief today (at least outside the actual world of business) is that the contract at will is so unfair and one-sided that it cannot be the outcome of a rational set of bargaining processes any more than, to take the extreme case, a contract for total slavery. While we may not, the criticism continues, be able to observe them, defects in capacity at contract formation nonetheless must be present: the ban upon the contract at will is an effective way to reach abuses that are pervasive but difficult to detect, so that modest government interference only strengthens the operation of market forces.[17]

In order to rebut this charge, it is necessary to do more than insist that individuals as a general matter know how to govern their own lives. It is also necessary to display the structural strengths of the contract at will that explain why rational people would enter into such a contract, if not all the time, then at least most of it. The implicit assumption in this argument is that contracts are typically for the mutual benefit of both parties. Yet it is hard to see what other assumption makes any sense in analyzing institutional arrangements (arguably in contradistinction to idiosyncratic, nonrepetitive transactions). To be sure, there are occasional cases of regret after the fact, especially after an infrequent, but costly, contingency comes to pass. There will be cases in which parties are naive, befuddled, or worse. Yet in framing either a rule of policy or a rule of construction, the focus cannot be on that biased set of cases in which the contract aborts and litigation ensues. Instead, attention must be directed to standard repetitive transactions, where the centralizing tendency powerfully promotes expected mutual gain. It is simply incredible to postulate that either employers or employees, motivated as they are by self-interest,

[17] Kronman, *Paternalism and the Law of Contracts*, 92 YALE L.J. 763, 777 (1983). The point is especially important in connection with the law of undue influence, where there is a long historical dispute over the relationship between the adequacy of consideration received and the procedural soundness of the underlying transaction. *See* Simpson, *The Horwitz Thesis and the History of Contracts*, 46 U. CHI. L. REV. 533, 561-80 (1979). Nonetheless, paternalistic explanations, whatever their force elsewhere, have little power in connection with employment relations. Indeed, if one thought it appropriate to restrict the powers of workers to make their own decisions during negotiations over the terms of employment, it might follow that restrictions on their right to participate in unions could be justified as well, for in both instances workers have proven that they often need to be protected against their own folly.

would enter routinely into a transaction that leaves them worse off than they were before, or even worse off than their next best alternative.

From this perspective, then, the task is to explain how and why the at-will contracting arrangement (in sharp contrast to slavery) typically works to the mutual advantage of the parties. Here, as is common in economic matters, it does not matter that the parties themselves often cannot articulate the reasons that render their judgment sound and breathe life into legal arrangements that are fragile in form but durable in practice.[18] The inquiry into mutual benefit in turn requires an examination of the full range of costs and benefits that arise from collaborative ventures. It is just at this point that the nineteenth-century view is superior to the emerging modern conception. The modern view tends to lay heavy emphasis on the need to control employer abuse. Yet, as the passage from *Payne* indicates,[19] the rights under the contract at will are fully bilateral, so that the employee can use the contract as a means to control the firm, just as the firm uses it to control the worker.

The issue for the parties, properly framed, is not how to minimize employer abuse, but rather how to maximize the gain from the relationship, which in part depends upon minimizing the sum of employer and employee abuse. Viewed in this way the private-contracting problem is far more complex. How does each party create incentives for the proper behavior of the other? How does each side insure against certain risks? How do both sides minimize the administrative costs of their contracting practices?

In order to show the interaction of all relevant factors, it is useful to analyze a case in which the problem of bilateral control exists, but where the overtones of inequality of bargaining power are absent. The treatment of partnership relations is therefore very instructive because partners are generally social and economic

---

[18] One striking example of the durability of at-will arrangements in other markets comes from an issue frequently litigated in the law of eminent domain: whether a tenant in possession under a lease is entitled to compensation for the simple expectation of renewal of the lease when the government condemns the fee or the leasehold itself. The legal conclusion is that the tenant should receive compensation equal to the market value of the leasehold that he could have received from a willing buyer, taking into account the expectation of renewal. *See* Almota Farmers Elevator & Warehouse Co. v. United States, 409 U.S. 470, 473-74 (1973). But the problem only arises because the value of the expectation to the tenant is high, which in turn suggests that the probability of renewal is great, and indeed often approaches one hundred percent.

[19] Payne v. Western & Atl. R.R., 81 Tenn. 507, 518-19 (1884). *See supra* note 3 and accompanying text.

equals between whom considerations of inequality of bargaining power, so evident in the debate over the contract at will, have no relevance. To be sure, the structural differences between partnership and employment contracts must be identified, but these will in the end explain why the at-will contract may make even greater sense in the employment context.

A. At-Will Arrangements in Partnerships

The economic motivation to form partnerships is easy enough to understand. Partnerships allow individuals to pool the capital necessary to undertake larger ventures, and they allow the parties to obtain the benefits of a division of labor within a single firm, where each partner retains some equity claim in the final output of the firm. By combining their separate resources and talents in a cooperative venture, the parties can produce more of value than they could by acting separately and then exchanging their outputs in discrete market transactions, such as sale or barter.

Nonetheless, the organization of the simplest firm, i.e., one with only two people, creates a set of problems that does not exist for the sole entrepreneur: the question of management. The efforts of the two individuals must be coordinated, and each person must be sure that some steps are taken to prevent the other party to the contract from acting to nullify or reduce the advantages that are promised at formation.

The first business question that confronts the partnership is the division of the proceeds. If these are to be evenly divided, then some steps, some costly steps, must be taken to ensure equal contributions. Thus, for example, the two parties may have different types of skills with different market values. Should some difference in the time allocated to the partnership be required? Should the less productive partner be required to commit some additional cash or property to the joint venture? Or should the equal division of partnership gains be abandoned in favor of some complex formula, with partners resorting to intricate capital structures complete with loans, preferred interests, and options to purchase? The matter is, moreover, not simply one of fairness when disputes occur, but also of the incentives created upon contract formation. The less effective the mechanics for controlling the contribution of resources and the distribution of gains, the less likely it will be that the venture will be formed, and the smaller will be the expected gains, both private and social.

A critical variable in partnership formation is the level of effort the parties will bring to the joint venture. One naive assump-

tion is that the cooperative nature of the enterprise will not influence the effort and performance levels of the partners. Yet, as the economic literature on agency costs so tellingly demonstrates,[20] hidden conflicts of interest pervade all business organizations at birth, during life, and upon death. The obvious conflict of interest arises when the gains of one partner are the losses of the other, as with simple theft of partnership assets by a partner. A second form of conflict, less dramatic but more persistent, arises because each partner bears the full costs of his own individual action, while obtaining only some fraction of the partnership gain that that action produces. The wedge necessarily driven between private and social (i.e., firm) costs creates an incentive for each partner to produce only to the point where, at the margin, the total amount of private gain is some multiple (greater than one) of the additional unit of private cost.[21] That tendency can manifest itself in any number of ways. Each partner may not work as hard as he would were he in business alone. Each partner may divert firm business to his own private account, all at a sporting discount, if only the customer remains quiet about the special arrangement. If both partners engage in this opportunistic behavior, then the firm output will fall below the levels it would achieve if they both continued to labor until, at the margin, *partnership* gains matched *partnership* costs.

These potential conflicts of interest will not materialize in each and every case, but they do pose a persistent threat to the stability of the firm. Within this framework, therefore, the managerial task is to determine what devices increase the likelihood that both partners will produce to the ideal point. In part there are legal prohibitions. The diversion of partnership assets can be met

---

[20] *See* Jensen & Meckling, *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, 3 J. FIN. ECON. 305, 308-10, 312-13, 333-34 (1976). For a recent collection of materials on the subject, see THE AGENCY RELATIONSHIP (J. Pratt & R. Zeckhauser eds. forthcoming), which contains a paper I wrote dealing with many of the same issues discussed here: Epstein, *Agency Cost, Employment Contracts, and Labor Unions*.

[21] The conflict can be stated formally as follows: let C represent costs to the individual, and G represent gains to the firm. Then assume any division of gain, p, for each partner that is strictly less than one. The conflict then arises whenever the following condition is established: $pG < C < G$. This inequality says that it is in the interest of the firm that the activity with cost C be undertaken, but that the individual actor will not undertake it because he is left worse off by the difference between C and pG. Good management rests in part upon the ability to narrow the gap between pG and G, or in the alternative to reduce the private costs to below pG. The challenge is to devise institutional arrangements that can overcome this fundamental conflict in discrete instances, and the problem in a sense is quite unending. If costs are reduced, then activities that were once unthinkable now become plausible, and with respect to some fraction of this new class, C will lie between pG and G.

with an action that recovers the proceeds for the benefit of the firm.[32] Yet litigation is always expensive and unpleasant. As a common sense matter an ounce of protection may be worth a pound of cure. With small partnerships one effective sanction is simply for each partner to watch the others: who comes to the office first, who leaves last; who does nonfirm work on firm time or with firm equipment; who receives the larger number of phone calls from customers; who generates the larger number of customer billings; whose services are in greater demand by outsiders.

The basic problem becomes only more complex as the size of the partnership increases—consider, for example, the division of profits in a large law firm, with its separate tiers of partners and, increasingly, permanent associates. A rough generalization is that as the size of the enterprise increases the demand for some internal organizational structure increases, perhaps exponentially, as well. Now the conflicts of interest become ever more acute as the gains from diversion increase because each partner receives a smaller fraction of firm profits.[33] Similarly, the informal social sanctions and affective ties among partners are attenuated by the sheer force of numbers. More formal procedures are required to control abuse, to protect, as it were, the well-intentioned partners from their fellows: personnel committees, formal audits, and managerial specialization quickly become standard parts of firm practice.

Partnership arrangements are difficult to police for yet another reason: all the partners are required to place all their eggs in

---

[32] The problem here is one of the misappropriation of either partnership or corporate opportunity, and has been recognized as an issue as early as Roman times:

> A partnership lasts as long as the parties remain of the same mind, but when one of them renounces the partnership, it is dissolved. But of course if one of the partners renounces for the purpose of profiting alone by some coming gain, for example, if my partner in a universal partnership [one embracing all assets, however acquired], having been left heir by somone, renounces the partnership in order to gain the inheritance for himself alone, he will be compelled to share this gain. If, however, he makes other gain which he has not sought for, this belongs to him alone. I, on the other hand, have the sole right to anything whatever that I acquire after his renunciation of the partnership.

GAIUS, INSTITUTES, III, 151 (F. De Zulueta trans. 1946). Note the at-will presumption coupled with rules to prevent diversions of opportunity from the partnership; these rules are skewed so that the party in breach is left worse off, since he must share his own private gains but has no share in the gains of the innocent party. The Roman rule is carefully crafted and in essence represents the modern law, though it is unlikely that the Romans had any understanding of the formal economic theory of agency.

[33] A more formal statement, using the symbols set out *supra* note 21, would begin by noting that as the size of the firm increases, p (the fraction of the gain that goes to each partner) decreases. Then, returning to the basic inequality ($pG < C < G$), we can see that as p decreases, the interval between pG and G increases, and with it the likelihood and severity of conflicts of interest.

a single basket. The principal contribution to most partnerships is the partner's labor. Labor, unlike money, is not easily divisible among multiple investment opportunities. Therefore, while partners obtain the many advantages of firm specialization, they must sacrifice the advantage of risk diversification that is normally available in capital markets.[34] Not only is it possible to diversify risk at low cost in capital markets, but it is also easy to redeploy assets across firms. The threat of such redeployment, in turn, may be the investor's most powerful means of protecting himself from abuse at the hands of the firm's managers, who are always wary of hostile tender offers.[35]

Labor markets differ from capital markets in both of these particulars. Labor cannot be diversified in the normal service partnership because a person cannot be in more than one or two separate ventures at any given time and hope to maintain productive activities. A mutual fund for jobs is quite unthinkable. In addition, entrance and exit in labor markets is highly complex, as is often stressed in the literature critical of the contract at will.[36] Thus, a party may be legally entitled to withdraw from a joint venture at will, but he cannot substitute someone else in his place without the consent of the remaining partners, given the delicate personal chemistry behind any joint venture.

The consequences of low diversification and impaired practical alienability are clear. Labor ventures are inherently more risky. The choice of partners is critical at the outset, as is the need to monitor their activities continually. Yet attending to these needs raises a powerful tension that pervades partnership relations. Long-term cooperative ventures require some permanent internal structure, while effective control against abuse often depends upon the ability to withdraw from the venture at any time, and without

---

[34] I have discussed some of these points in a similar vein in Epstein, *supra* note 20. For a convenient explanation of the principles of portfolio diversification, see Langbein & Posner, *Market Funds and Trust-Investment Law* (pt. 1), 1976 AM. B. FOUND. RESEARCH J. 1; *see also* Langbein & Posner, *Market Funds and Trust-Investment Law* (pt. 2), 1977 AM. B. FOUND. RESEARCH J. 1.

[35] Because the costs to shareholders of monitoring managerial decisions are high, they find it in their self-interest to remain passive and to avoid losses by simply selling their shares. *See* Easterbrook & Fischel, *The Proper Role of a Target's Management in Responding to a Tender Offer*, 94 HARV. L. REV. 1161, 1170-71 (1981).

[36] *See, e.g.*, Blades, *supra* note 4, at 1405 ("Obviously, if every employee could go from job to job with complete ease, there would be little need to provide other means of escape from the improper exertion of employer pressure."). The effect of impaired mobility upon the desirability of the contract at will is discussed *infra* notes 45-46 and accompanying text.

cause, as most partnership agreements provide.[27]

The logic behind the right of withdrawal depends upon a delicate balance of advantage. In the simplest case, assume that the partnership originally calls for an even division of the gains between two partners. If partner *A* knows (or even has reason to believe) that partner *B* has misappropriated assets or opportunities belonging to the firm, his threat to withdraw will become instantly more credible because his losses from withdrawal must be measured against a (revised) baseline of less than half the partnership gains. *B*'s expected losses from dissolution will be correspondingly greater because he will lose the expected illicit gains obtained from improper behavior. *A*'s threat tends therefore to be made under those circumstances where it is most called for.

The at-will provision in partnership agreements by no means controls all abuse—nothing does that—because both sides will have to bear the costs of dissolution (including a costly partnership accounting) in the event that a break-up occurs. It is quite possible, for example, for *B* to engage in activities that net him more than his agreed-upon share of the assets if he knows that *A*'s costs of breaking up exceed his expected gains from dissolution. Nonetheless, the greater the wrongdoing by *B*, the more likely it is that dissolution will cost *A* less than continued abuse by *B*. The sanction of withdrawal remains imperfect, but it tends to become more credible as the abuses become greater. The provision for at-will dissolution of the partnership helps stabilize the arrangement after formation and thus tends to be in the interest of both parties at formation. To be sure, explicit legal remedies may be desired in cases of waste, but direct legal action always costs more than the simple act of separation, and any lawsuit is costly to bring and uncertain in its outcome. Thus, while resort to legal remedies is to be expected, it will not be frequent. Self-help by withdrawal may have a lower payoff, but its lower cost and greater predictability make it an effective sanction against misbehavior.

B. Employment at Will

When we move from partnerships to employment relations, we must determine how the differences in the relationships between

---

[27] *See* Hillman, *Misconduct as a Basis for Excluding or Expelling a Partner: Effecting Commercial Divorce and Securing Custody of the Business*, 78 NW. U.L. REV. 527, 531 (1983); Hillman, *The Dissatisfied Participant in the Solvent Business Venture: A Consideration of the Relative Permanence of Partnerships and Close Corporations*, 67 MINN. L. REV. 1, 4 (1983).

the parties influence the mix of the formal and informal control devices available to the parties. The starting point of the analysis is the difference in the ways in which profit and loss are divided. The central feature of a partnership is that the residual gains and losses are shared by the partners under an internal formula that makes both parties equity claimants. In the employment-contracting situation, the employer is the sole residual claimant upon the earnings of the firm, while the employee receives a fixed wage.[28] But this important difference does not mean that the advantages of the at-will arrangement are of no importance to the employment relationship. On the contrary, it is possible to identify a number of reasons why the at-will contract usually works for the benefit of both sides in employment as well as partnership contexts.

1. *Monitoring Behavior.* The shift in the internal structure of the firm from a partnership to an employment relation eliminates neither bilateral opportunism nor the conflicts of interest between employer and employee. Begin for the moment with the fears of the firm, for it is the firm's right to maintain at-will power that is now being called into question. In all too many cases, the firm must contend with the recurrent problem of employee theft and with the related problems of unauthorized use of firm equipment and employee kickback arrangements.[29] As the analysis of partnerships shows, however, the proper concerns of the firm are not limited to obvious forms of criminal misconduct. The employee on a fixed wage can, at the margin, capture only a portion of the gain from his labor, and therefore has a tendency to reduce output.[30] The employee who receives a commission equal to half the

---

[28] Sometimes the employee's wage will be fixed, not in dollar terms, but as a function of his own productivity (for example, sales personnel who work on commission). But even if the employee's claim is a function of the firm's profit from particular transactions, he is in a different position than is an owner. An individual salesman can make a fortune while the firm loses money; a pure equity partner cannot.

[29] For example, in Perks v. Firestone Tire & Rubber Co., 611 F.2d 1363, 1364 (3d Cir. 1979), the plaintiff, a discharged employee, had allowed the representative of a supplier to procure prostitutes for him.

[30] By way of comparison, his position is like that of a lienholder who is quite happy so long as the value of the security remains above the level of the lien, even if the venture itself does not prosper. This is one reason why there is an enormous reluctance to allow the mortgagee to enter into possession before default, especially when the equity in the property is large. The risks, of course, change radically upon default, at which point a mortgagee will generally want to keep up the property because its value is less than that of his lien. Note too that the conflict of interest is of great importance to an equity investor in a limited partnership who, having relied upon the valuation of property prepared for the mortgage lender, discovers that the lender's assessor was less sensitive than he is to the positive po-

firm's profit attributable to his labor may work hard, but probably not quite as hard as he would if he received the entire profit from the completed sale, an arrangement that would solve the agency-cost problem only by undoing the firm.

The mismatch between benefits received and capital or labor contributed, then, can and does exist between employer and employee just as it does between partner and partner. The different ways in which firm proceeds are divided only determines where and how the conflicts will arise, not whether they will arise. Indeed, since the roles of employer and employee are radically asymmetrical, the potential source of conflict is, if anything, larger than it is in relations between copartners. The conflicts between employer and employee may sometimes call for the severance of the relationship, just as they do in the partnership context. But the rational response is to counteract the tendency for employee abuse only to the point where private gain equals private cost.[31] Agency costs are like other costs that must be minimized in order for production to proceed, and the persistence of firms shows that this can be done.

The problem of management then is to identify the forms of social control that are best able to minimize these agency costs. Here the choices are the same as they were in the partnership situation. One obvious form of control is the force of law. The state can be brought in to punish cases of embezzlement or fraud. But this mode of control requires extensive cooperation with public officials and may well be frustrated by the need to prove the criminal offense (including mens rea) beyond a reasonable doubt, so that vast amounts of abuse will go unchecked. Private litigation instituted by the firm may well be used in cases of major grievances, either to recover the property that has been misappropriated or to prevent the individual employee from further diverting firm busi-

---

tential of the project.

[31] There are analogies with corporate financial structures. Thus there is always a partial conflict of interest when a firm contains both debt and equity in its capital structure. The holders of the equity are, by virtue of the existence of the debt, induced to engage in riskier ventures than they would undertake without the debt. The explanation is clear enough. With the debt, the equity holders still obtain all the profits from the firm. But now the risk of bankruptcy is shared by others, thereby allowing the partial externalization of the costs of failure. *See* De Alessi, *Private Property and Dispersion of Ownership in Large Corporations*, 28 J. FIN. 839 (1973). The point does not condemn such debt structures, but it does invite the use of other control measures, such as limitations, contained in the original loan agreement, on the types of activities that the borrower may undertake. *See generally* Smith & Warner, *On Financial Contracting: An Analysis of Bond Covenants*, 7 J. FIN. ECON. 117 (1979).

ness to his own account. But private litigation, like public prosecution, is too blunt an instrument to counter employee shirking or the minor but persistent use of firm assets for private business.

As with the partnership, some alternative forms of control are needed. Internal auditors may help control some forms of abuse, and simple observation by coworkers may well monitor employee activities. (There are some very subtle tradeoffs to be considered when the firm decides whether to use partitions or separate offices for its employees.) Promotions, bonuses, and wages are also critical in shaping the level of employee performance.[32] But the carrot cannot be used to the exclusion of the stick. In order to maintain internal discipline, the firm may have to resort to sanctions against individual employees. It is far easier to use those powers that can be unilaterally exercised: to fire, to demote, to withhold wages, or to reprimand. These devices can visit very powerful losses upon individual employees without the need to resort to legal action, and they permit the firm to monitor employee performance continually in order to identify both strong and weak workers and to compensate them accordingly. The principles here are constant, whether we speak of senior officials or lowly subordinates, and it is for just this reason that the contract at will is found at all levels in private markets.

The parallels to the partnership situation are instructive, for the at-will arrangement is, if anything, even more effective between employers and employees. As with partnerships, the threat, be it of discharge or resignation, becomes more effective the greater the level of employee or employer abuse; it is thus an effective if informal means of encouraging self-restraint. In addition, within the employment context firing does not require a disruption of firm operations, much less an expensive division of its assets. It is instead a clean break with consequences that are immediately clear to both sides. The lower cost of both firing and quitting, therefore, helps account for the very widespread popularity of employment-at-will contracts. There is no need to resort to any theory of economic domination or inequality of bargaining power to explain at-will contracting, which appears with the same tenacity in relations

---

[32] *See, e.g.*, Malcomson, *Work Incentives, Hierarchy, and Internal Labor Markets*, 92 J. POL. ECON. 486 (1984), for one effort to model the effectiveness of using internal promotions to improve workers' performance when monitoring is costly and hence imperfect. Malcomson's model does not address the question of whether the workers are under term contracts or contracts at will, but the structure of his argument is consistent with the at-will model of legal relations.

between economic equals and subordinates and is found in many complex commercial arrangements, including franchise agreements, except where limited by statutes.[33]

Thus far, the analysis generally has focused on the position of the employer. Yet for the contract at will to be adopted ex ante, it must work for the benefit of workers as well. And indeed it does, for the contract at will also contains powerful limitations on employers' abuses of power. To see the importance of the contract at will to the employee, it is useful to distinguish between two cases. In the first, the employer pays a fixed sum of money to the worker and is then free to demand of the employee whatever services he wants for some fixed period of time. In the second case, there is no fixed period of employment. The employer is free to demand whatever he wants of the employee, who in turn is free to withdraw for good reason, bad reason, or no reason at all.

The first arrangement invites abuse by the employer, who can now make enormous demands upon the worker without having to take into account either the worker's disutility during the period of service or the value of the worker's labor at contract termination. A fixed-period contract that leaves the worker's obligations unspecified thereby creates a sharp tension between the parties, since the employer receives all the marginal benefits and the employee bears all the marginal costs.[34]

Matters are very different where the employer makes increased demands under a contract at will. Now the worker can quit whenever the net value of the employment contract turns negative. As with the employer's power to fire or demote, the threat to quit

---

[33] As an example of a restriction upon the power of termination, automobile dealers may recover damages resulting from a manufacturer's failure to act in good faith in not renewing the dealer's franchise. 15 U.S.C. §§ 1221-1225 (1982). These provisions were conceived of as a supplement to the antitrust laws. Act of Aug. 8, 1956, Pub. L. No. 1026, 70 Stat. 1125 (codified at 15 U.S.C. §§ 1221-1225 (1982)) (statement of purpose).

[34] This makes it difficult to accept the argument that "[e]mployment at will is the ultimate guarantor of the capitalist's authority over the worker." Feinman, *supra* note 4, at 132-33. Yet, as Feinman notes, historically the employees who brought actions were generally those who earned high salaries. *Id.* at 118, 131-33. This is simply enough explained by noting that those would have been the only cases in which the amount in controversy exceeded the plaintiff's expected costs of suit. But the fact that employees could also quit at will makes it difficult to see in the at-will device the exploitation of the working class, especially since real wages were continually rising throughout the eighteenth and nineteenth centuries. *See* D. DIAMOND & J. GUILJOIL, UNITED STATES ECONOMIC HISTORY 277, 468 (1973); HERMAN E. KROOSS, AMERICAN ECONOMIC DEVELOPMENT 395-98 (2d ed. 1966); S. RATNER, J. SOLTOW & R. SYLLA, THE EVOLUTION OF THE AMERICAN ECONOMY 247, 308-09 (1979). Nor does a theory focusing on the employer's authority account for the right of the employee to *quit* at will.

(or at a lower level to come late or leave early) is one that can be exercised without resort to litigation. Furthermore, that threat turns out to be most effective when the employer's opportunistic behavior is the greatest because the situation is one in which the worker has least to lose. To be sure, the worker will not necessarily make a threat whenever the employer insists that the worker accept a less favorable set of contractual terms, for sometimes the changes may be accepted as an uneventful adjustment in the total compensation level attributable to a change in the market price of labor. This point counts, however, only as an additional strength of the contract at will, which allows for small adjustments *in both directions* in ongoing contractual arrangements with a minimum of bother and confusion.

The case for the contract at will is further strengthened by another feature common to contracts of this sort. The employer is often required either to give notice or to pay damages in lieu of notice; damages are traditionally equal to the wages that the employee would have earned during the notice period. These provisions for "severance pay" provide the worker with some protection against casual or hasty discharges, but they do not interfere with the powerful efficiency characteristics of the contract at will. First, lump-sum transfers do not require the introduction of any "for cause" requirement, which could be the source of expensive litigation. Second, because the sums are definite, they can be easily computed, so that administrative costs are minimized. Third, because the payments are unconditional, they do not create perverse incentives for the employee or heavy monitoring costs for the employer: the terminated employee will not be tempted to avoid gainful employment in order to run up his damages for wrongful discharge; the employer, for his part, will not have to monitor the post-termination behavior of the employee in order to guard against that very risk. Thus, provisions for severance pay can be used to give employees added protection against arbitrary discharge without sacrificing the advantages of a clean break between the parties.

2. *Reputational Losses.* Another reason why employees are often willing to enter into at-will employment contracts stems from the asymmetry of reputational losses. Any party who cheats may well obtain a bad reputation that will induce others to avoid dealing with him. The size of these losses tends to differ systematically between employers and employees—to the advantage of the employee. Thus in the usual situation there are many workers and a single employer. The disparity in number is apt to be greatest in

large industrial concerns, where the at-will contract is commonly, if *mistakenly*, thought to be most unsatisfactory because of the supposed inequality of bargaining power.[35] The employer who decides to act for bad reason or no reason at all may not face any legal liability under the classical common law rule. But he faces very powerful adverse economic consequences. If coworkers perceive the dismissal as arbitrary, they will take fresh stock of their own prospects, for they can no longer be certain that their faithful performance will ensure their security and advancement. The uncertain prospects created by arbitrary employer behavior is functionally indistinguishable from a reduction in wages unilaterally imposed by the employer. At the margin some workers will look elsewhere, and typically the best workers will have the greatest opportunities.[36] By the same token the large employer has more to gain if he dismisses undesirable employees, for this ordinarily acts as an implicit increase in wages to the other employees, who are no longer burdened with uncooperative or obtuse coworkers.

The existence of both positive and negative reputational effects is thus brought back to bear on the employer. The law may tolerate arbitrary behavior, but private pressures effectively limit its scope. Inferior employers will be at a perpetual competitive disadvantage with enlightened ones and will continue to lose in market share and hence in relative social importance. The lack of legal protection to the employees is therefore in part explained by the increased informal protections that they obtain by working in large concerns.[37]

3. *Risk Diversification and Imperfect Information.* The

---

[35] *See supra* note 5 and accompanying text. The possibility of the poor exploiting the rich is discussed in the context of the free-rider problem by Mancur Olson, who notes that it is often easy for a small property owner, for example, to free-ride off his larger neighbor's provision of certain common goods. MANCUR OLSON, THE LOGIC OF COLLECTIVE ACTION 27-29 (1965).

[36] To illustrate the problem, consider the firm that wants to reduce the size of its work force. If it makes a general promise of ample severance pay to whoever leaves the firm, it runs the very high risk that it will lose its most desirable employees, for it is just these workers who will find it easiest to recoup their income elsewhere. There are complications, however, since the skills of able employees may not be transferable, or at least may not be easily marketable. But the tendency is nonetheless clearly present.

[37] The point has a close analogue in the law of landlord and tenant. Landlords of large apartment complexes are often able to extract more favorable terms from tenants than the owner of a downstairs duplex can. But the large landlord has more to lose by capricious behavior, for he suffers from the same asymmetric reputational vulnerability as does the large employer. The greater legal power conferred on him by contract thus works to the benefit of both sides, as any tenant who has been relieved by the eviction of a noisy neighbor can easily attest.

contract at will also helps workers deal with the problem of risk diversification. In this regard the employee is in the same position as the partner. Ordinarily, employees cannot work more than one, or perhaps two, jobs at the same time. Thereafter the level of performance falls dramatically, so that diversification brings in its wake a low return on labor. The contract at will is designed in part to offset the concentration of individual investment in a single job by allowing diversification among employers *over time*. The employee is not locked into an unfortunate contract if he finds better opportunities elsewhere or if he detects some weakness in the internal structure of the firm. A similar analysis applies on the employer's side where he is a sole proprietor, though ordinary diversification is possible when ownership of the firm is widely held in publicly traded shares.

The contract at will is also a sensible private adaptation to the problem of imperfect information over time. In sharp contrast to the purchase of standard goods, an inspection of the job before acceptance is far less likely to guarantee its quality thereafter. The future is not clearly known. More important, employees, like employers, *know what they do not know*. They are not faced with a bolt from the blue, with an "unknown unknown." Rather they face a known unknown for which they can plan. The at-will contract is an essential part of that planning because it allows both sides to take a wait-and-see attitude to their relationship so that new and more accurate choices can be made on the strength of improved information. ("You can start Tuesday and we'll see how the job works out" is a highly intelligent response to uncertainty.) To be sure, employment relationships are more personal and hence often stormier than those that exist in financial markets, but that is no warrant for replacing the contract at will with a for-cause contract provision. The proper question is: will the shift in methods of control work a change for the benefit of both parties, or will it only make a difficult situation worse?[88]

---

[88] We should greet with skepticism any claim that takes the following form:

> The at will doctrine should be altered not because of "unequal bargaining power," but rather because it is inefficient. Courts must intervene, according to this view, in order to bring about the substantive outcome that the parties would have reached had transaction and information costs not precluded informed negotiation. When high costs of bargaining prevent negotiation between individual employees and employers, and inadequate access to information prevents parties from properly valuing the benefits of job security, judicial intervention is justified to ensure a more efficient result.

Note, *Wrongful Discharge*, *supra* note 4, at 1830. The author is right to dismiss inequality of bargaining power as a makeweight argument. But the discussion of imperfect information is nothing short of mystifying, for it simply assumes that universal arrangements are univer-

4. *Administrative Costs.* There is one last way in which the contract at will has an enormous advantage over its rivals. It is very cheap to administer. Any effort to use a for-cause rule will in principle allow all, or at least a substantial fraction of, dismissals to generate litigation. Because motive will be a critical element in these cases, the chances of either side obtaining summary judgment will be negligible. Similarly, the broad modern rules of discovery will allow exploration into every aspect of the employment relation. Indeed, a little imagination will allow the plaintiff's lawyer to delve into the general employment policies of the firm, the treatment of similar cases, and a review of the individual file. The employer for his part will be able to examine every aspect of the employee's performance and personal life in order to bolster the case for dismissal.

Nonetheless, it may be said that this inquiry is worth conducting because employers err in making decisions to fire and injustices will be done unless legal sanctions are imposed. But this analysis entirely ignores the fact that error costs always run in both directions. It has already been shown that there are powerful correctives against capricious discharge even under an at-will rule.[39] The chances of finding an innocent employee wronged by a firm vendetta are quite remote. By the same token, jury sympathy with aggrieved plaintiffs may result in a very large number of erroneous verdicts for employees. In principle it might be proper to tolerate the high error rate if the consequences of erroneous dismissal to the innocent employee were more severe than the consequences of erroneous reinstatement to the innocent employer. But quite the opposite is apt to be the case. Able employees are the very persons who have the greatest opportunity of obtaining alternate employment in the marketplace and who can therefore best mitigate their losses. Although their search for new work may be complicated because of the previous dismissal, the dismissed employee usually can get other persons, e.g., representatives of other companies with whom he has dealt, to help overcome the negative inference from the dismissal. Indeed there is less trouble in explaining away the dismissal if it is generally understood that contracts are terminable at will, since termination no longer implies employee misconduct.[40]

---

sally unsound, without recognizing any of their strengths or noting any of the defects in the alternative rules.

[39] *See supra* notes 36-37 and accompanying text.

[40] Other (imperfect) legal protections are available to the employee as well. If, for ex-

The results of this analysis are not upset in any way by the procession of cases that now reach the appellate courts.[41] First, these cases are never a random sample of all dismissals, but rather are selected because they promise the greatest chance of success under the applicable law. Yet the doctrines of wrongful discharge apply to all cases, including those in which superior results are reached by a simple application of the at-will rule. In addition, these cases frequently arise on a motion to dismiss, so that their assertions of improper motive may well be left unsubstantiated at trial. Finally, there is good reason for skepticism about the power of juries to divine motive and purpose from the evidence that is presented to them. A single case easily can be regarded either as employer oppression or employer benevolence, and there is every reason to expect that very different interpretations of similar fact patterns will proliferate under any version of the for-cause standard.[42]

---

ample, the previous employer deliberately gives false information in response to inquiries from prospective employers, his conduct may be actionable under the law of defamation, since the employer's bad faith would overcome any qualified privilege the employer might enjoy. If the employer volunteers the false information, the standard of liability may well be even stricter. *See, e.g.*, Clark v. Molyneux, 3 Q.B.D. 237, 243-44 (1877); R. Epstein, C. Gregory & H. Kalven, Cases and Materials on Torts 1154-57 (4th ed. 1984); *see also* W. Page Keeton, Prosser and Keeton on the Law of Torts § 115, at 832-34 (5th ed. 1984).

[41] *See* cases cited *supra* notes 6-8.

[42] Thus Monge v. Beebe Rubber Co., 114 N.H. 130, 316 A.2d 549 (1974), lends grist to the skeptic's mill. The facts of the case were summarized by another court as follows: "[f]emale employee wrongfully discharged for refusing to date foreman; court rejected employment-at-will defense as 'based on ancient feudal system.' " Novosel v. Nationwide Ins. Co., 721 F.2d 894, 897 n.5 (3d Cir. 1983). The reference to the feudal system is amusing since the contract at will by all accounts came to the fore in the last half of the nineteenth century, in part as a response to industrialization. *See* Feinman, *supra* note 4, at 125-29.

The summary of the facts found in the dissent to *Monge* itself, however, throws a very different light on the case:

> In my view, reasonable men could not find for the plaintiff on the evidence in this case even under the new rule of law which the court has fashioned today. The substance of the plaintiff's claim is that she was discharged because she did not accept an invitation of her foreman to go out with him. Although it was denied by the foreman, the jury could find on plaintiff's testimony alone that the invitation was extended. It was a single instance, however, and there is no claim that it was repeated or further pursued. It is not findable that this single refusal was the reason for the termination of plaintiff's employment. There was evidence, and none to the contrary, that it was a shortage of work and her lowest seniority that caused her press machine to be shut down and her loss of overtime. When her machine was shut down, she was given work on a degreasing machine at a higher rate of pay than when she started. When she told the foreman she "needed the money" from the overtime, he offered what from the uncontradicted evidence was the only work available to help her out until her overtime was restored. The only so-called harassment and ridicule claimed amounts to no more than once saying "How do you like my floor boy?" and "My wife wouldn't do that." It is uncontradicted that when she was having trouble with annoying phone calls and needed help, the per-

The difficulties are even greater once it becomes established that dismissals cannot be made at will. Now the employer will have to reconsider every aspect of personnel relations. If it is improper to dismiss at will, then it becomes improper to demote or to transfer at will, for an employee will be able to assert with perfect propriety that the employer had made work so unattractive to him that his conduct should amount to a "constructive dismissal" for which either damages or reinstatement is appropriate. Thus a rule that starts with modest ambitions will in the end regulate each and every aspect of the employment relationship. Professor Clyde Summers, for example, in his own proposal for the creation of a statutory cause of action for unjust dismissal recognizes that "[t]he statute [on unjust dismissal] must reach all forms of disciplinary action related to an employee's job, including demotion, reduction in pay, reduction in seniority, assignment to undesirable work, and forced resignation."[48] But he does not explain why the necessity for fashioning comprehensive regulation counts as a virtue instead of a vice, in light of the deleterious effects of increased regulation upon expected hiring patterns of employers. Where an employer might have been more willing to take risky employees under an at-will rule, he will now be less willing to do so under the for-cause rule because any subsequent demotion or dismissal will be an open invitation to a lawsuit by an aggrieved employee. Furthermore, in most at-will situations the dismissed employee is replaced by another, so it is hard to see how employees *as a class* benefit from a rule that can only hamper general mobility in labor markets.

These difficulties arise, moreover, no matter what the form of the rule. Thus, the rule could be stated as one that prohibits dismissals (or transfers) without cause, or it could be phrased only to prohibit a dismissal that is made in bad faith or with malice. The differences in formal expression will undoubtedly be significant in

---

sonnel manager personally went to the police and then to her home to talk with her and her husband; that when she could not pick up her Christmas turkey, the foreman personally delivered two instead of one to her home; and that he also at her request gave her husband, a mechanic, work on his automobile.

Her final termination was in accordance with established company rules and she neither contested the termination nor pursued the grievance procedures under the union contract. She was denied unemployment compensation on the ground that she was a "voluntary quit" and did not appeal that finding.

A finding that this company discharged the plaintiff because she refused her foreman a date eight months before could not reasonably be made and should not be permitted to stand.

Monge, 114 N.H. at 134-35, 316 A.2d at 552-53 (dissenting opinion).

[48] Summers, *supra* note 4, at 526-27.

the litigation of individual cases, as the for-cause standard should impose more stringent restrictions on employer freedom than the bad-faith standard. But from the institutional point of view the differences between the types of regulatory schemes would have only minor impact because *all* of them place costly and inconvenient restraints upon contractual freedom. So long as the cases align themselves in any continuous array, a large number of dismissals will be fair game for litigation no matter what the applicable legal standard. The administrative and error-cost considerations therefore reinforce the basic conclusion: there is no warrant for legal interference with the contract at will.

5. *Bilateral Monopoly and Inequality of Bargaining Power.* The account thus far given of the contract at will in no way depends upon any notion of an inherent inequality of bargaining power that pervades all employment contracts. Indeed if such an inequality did govern the employment relationship, we should expect to see conditions that exist in no labor market. Wages should be driven to zero, for no matter what their previous level, the employer could use his (inexhaustible) bargaining power to reduce them further, until the zero level was reached.[44] Similarly, inequality of bargaining power implies that the employee will be bound for a term while the employer (who can pay the peppercorn consideration) retains the power to terminate at will. Yet in practice we observe both positive wages and employees with the right to quit at will.

The reason why these contracts at will are effective is precisely that the employer must always pay an implicit price when he exercises his right to fire. He no longer has the right to compel the employee's service, as the employee can enter the market to find another job. The costs of the employer's decision therefore are borne in large measure by the employer himself, creating an implicit system of coinsurance between employer and employee against employer abuse. Nor, it must be stressed, are the costs to the employer light. It is true that employees who work within a firm acquire specific knowledge about its operation and upon dismissal can transfer only a portion of that knowledge to the new

[44] Note that the same arguments could be made in the consumer market as well, leading to the prediction that these same workers with a zero wage will pay infinite prices for the necessities of life. The only coherent models are those that assume that total wages and other income supply a budget constraint for purchases, so that both markets are in equilibrium simultaneously. Abstracting one market from another is a ploy that makes inequality of bargaining power seem more plausible than it is.

job.[48] Nonetheless, the problem is roughly symmetrical, as the employer must find, select, and train a replacement worker who may not turn out to be better than the first employee. Workers are not fungible, and sorting them out may be difficult: resumes can be misleading, if not fraudulent; references may be only too eager to unload an unsuitable employee; training is expensive; and the new worker may not like the job or may be forced to move out of town. In any case, firms must bear the costs of voluntary turnover by workers who quit, which gives them a frequent reminder of the need to avoid self-inflicted losses. The institutional stability of employment contracts at will can now be explained in part by their legal fragility. The right to fire is exercised only infrequently because the threat of firing is effective.

Thus far the account of inequality of bargaining power has been wholly negative. But the description of the employment relationship does suggest one way in which inequality can arise, even within the framework of generally competitive markets. In the course of an ongoing relationship between employee and employer, each side gains from the contract more than it could obtain by returning to the open market. The surplus that is created must be divided between the parties. In principle, either the worker or the employer could receive the entire surplus without inducing the other party, who still receives a competitive return, to sever the relationship. A fortiori any solution that divides the surplus between the parties should be stable as well. The contract at will thus creates a bilateral monopoly, but only to the extent of the surplus.

The question of inequality of bargaining power can now be helpfully restated: which side will appropriate most of the surplus in any negotiations between them? Unlike the typical formulations of the problem, this leaves the set of possible solutions strictly bounded because the employee cannot be driven below the competitive wage and the employer cannot be driven to a wage above the sum of the competitive wage plus the full amount of the surplus. An employer can therefore be said to possess an inequality of bargaining power when he is able to appropriate more than half the surplus, while the employee can be said to possess inequality of bargaining power if *he* can appropriate more than half the surplus.

---

[48] In some cases it is all too transferable, as with customer lists or trade secrets. When such information is transferred, an employer may look for redress to the elaborate body of law that regulates the transmission of trade secrets. *See* Kitch, *The Law and Economics of Rights in Valuable Information*, 9 J. LEGAL STUD. 683, 689-701 (1980).

To take an example, assume the employer is prepared to pay 20, while the worker is willing to work for 10. The agreed wage therefore could fall anywhere between those two numbers. If the employer is systematically able to appropriate more than 5 of this surplus, by keeping the wage level below 15, then he has unequal bargaining power, though still within the framework of overall competitive markets.

The existence of some surplus should be pervasive in all labor markets, given that labor is not perfectly fungible. In practice, the size of the surplus on average should be relatively small at the time of contract formation. Because the parties have not built up much specific capital in the relationship, quitting or firing will cause relatively small dislocations. As time passes, however, the gains to both sides from continuing the employment relationship are apt to increase, so that both sides have more to lose from separation. The bilateral monopoly problem now assumes greater significance. The increased size of the surplus can easily make wages somewhat indeterminate (which is why workers are commonly nervous about asking for a raise, and employers are nervous about refusing it). As the stakes become larger, the amount of resources spent in obtaining a larger portion of the surplus should increase. A contractual breakdown should nonetheless be an infrequent occurrence, as both sides have strong incentives to keep the relationship viable. The costs of negotiation tend to be reduced because each side is familiar with the other. The scope for bluffing is somewhat limited by each party's knowledge of the preferences of the other side. Finally, there are strong reasons for each side to avoid squeezing the last drop out of a relationship: miscalculation of the reserve price of the other party (i.e., the minimum he will accept or the maximum he will pay) could lead to a severance of the relationship and thus to a loss of the entire surplus.

It still remains to be determined which side is likely to appropriate most of this contract-specific surplus. One might guess that the employer will be able to achieve this objective, perhaps because his experience in repeat transactions with many workers fosters greater skills in negotiation. In addition, the employer may know in general the market wages available to beginning workers, as these typically will be public knowledge. Yet a number of considerations suggest the opposite conclusion. First, the employer often bargains through subordinate managers and thus faces an agency-cost problem avoided by the worker who bargains on his own account. Second, the worker's opportunity cost for his time will often be lower than the employer's, so that the increased time he can

spend on the transaction may offset the employer's greater skill, if any, per unit of time. Third, the worker may be able to learn something about the employer's reservation price (i.e., the maximum wage he would be willing to pay) because the employer must reveal some information about his willingness to pay in negotiations with other long-term workers. Finally, it is not clear that the employer gains any real advantage because of his greater relative wealth, if any. To be sure, the wealthy employer can hold out for a larger share of the surplus because he has less, proportionally, to lose. Yet by the same token the employer's resolve may be weaker because he has less to gain by holding out.

This modest catalogue of considerations shows how difficult it is to determine the exact division of the surplus, although my suspicion is that in the broad run of cases it will tend to be evenly divided. But even if this guess is wrong, there is no reason for the law to interfere in the bargaining process. The whole question of inequality of bargaining power arises in the bounded context of how much of a *supra*competitive wage the worker will obtain. At the very worst, the worker will get the amount that is offered in some alternate employment where he has built up no specific capital. To try to formulate and administer a set of legal rules that will allow some trier of fact to measure the size of the surplus embedded in the ongoing transaction, and to allocate half (or more) of it to the worker, cannot be done at any social cost that is less than the expected size of the surplus itself, if it can be done at all. The entire exercise is fraught with the possibility of real error, as real resources would have to be expended solely to make transfer payments that can in no way enhance productive efforts. The existence of this transactional surplus does not negate the fact that markets are still competitive before prospective employers and employees enter into any transaction at all.

The size of the surplus, and thus the scope of any inequality problem, can be reduced more effectively by adopting legal rules that remove or minimize legal impediments to labor mobility. The contract at will, by allowing either side to sever relationships without legal impediment, tends to reduce rigidities in markets and thus to act as a counterweight to the bilateral monopoly problem that emerges even in voluntary markets. The complex rules that give workers "property" rights in their jobs tend to increase the size of any possible surplus and exacerbate the basic problem. The identification of a transaction-specific surplus, then, adds to our understanding of long-term employment relationships, but it affords no warrant for upsetting the contract at will on supposed

grounds of public policy.

III. Distributional Concerns

Enough has been said to show that there is no principled reason of fairness or utility to disturb the common law's longstanding presumption in favor of the contract at will. It remains to be asked whether there are some hitherto unmentioned distributional consequences sufficient to throw that conclusion into doubt. One clear sign that there are not is that the advocates of the wrongful-discharge action themselves have not cast the argument in this form. Professor Lawrence Blades, for example, makes clear from the title of his own paper—*Employment at Will v. Individual Freedom*[46]—that he thinks abrogation of the contract at will helps advance the cause of individual liberty. While I believe that he is wrong in that conclusion, I think that he chose the correct ground on which to fight, for a moment's reflection makes it clear that distributional considerations enter at best only obliquely into the employment context.

The proposed reforms in the at-will doctrine cannot hope to transfer wealth systematically from rich to poor on the model of comprehensive systems of taxation or welfare benefits.[47] Indeed it is very difficult to identify in advance any deserving group of recipients that stands to gain unambiguously from the universal abrogation of the at-will contract. The proposed rules cover the whole range from senior executives to manual labor. At every wage level, there is presumably some differential in workers' output. Those who tend to slack off seem on balance to be most vulnerable to dismissal under the at-will rule; yet it is very hard to imagine why some special concession should be made in their favor at the expense of their more diligent fellow workers.

The distributional issues, moreover, become further clouded once it is recognized that any individual employee will have interests on both sides of the employment relation. Individual workers participate heavily in pension plans, where the value of the holdings depends in part upon the efficiency of the legal rules that govern the companies in which they own shares. If the regulation of the contract at will diminishes the overall level of wealth, the losses are apt to be spread far and wide, which makes it doubtful that there are any gains to the worst off in society that justify

---

[46] Blades, *supra* note 4.

[47] I have addressed these issues in detail in Epstein, *supra* note 3.

somewhat greater losses to those who are better off. The usual concern with maldistribution gives us situations in which one person has one hundred while each of one hundred has one and asks us to compare that distribution with an even distribution of, say, two per person. But the stark form of the numerical example does not explain how the skewed distribution is tied to the concrete choice between different rules governing employment relations. Set in this concrete context, the choices about the proposed new regulation of the employment contract do not set the one against the many but set the many against each other, all in the context of a shrinking overall pie. The possible gains from redistribution, even on the most favorable of assumptions about the diminishing marginal utility of money, are simply not present.

If this is the case, one puzzle still remains: who should be in favor of the proposed legislation? One possibility is that support for the change in common law rules rests largely on ideological and political grounds, so that the legislation has the public support of persons who may well be hurt by it in their private capacities.[48] Another possible explanation could identify the hand of interest-group politics in some subtle form. For example, the lawyers and government officials called upon to administer the new legislation may expect to obtain increased income and power, although this explanation seems insufficient to account for the current pressure. A more uncertain line of inquiry could ask whether labor unions stand to benefit from the creation of a cause of action for wrongful discharge. Unions, after all, have some skill in working with for-cause contracts under the labor statutes that prohibit firing for union activities, and they might be able to promote their own growth by selling their services to the presently nonunionized sector. In addition, the for-cause rule might give employers one less reason to resist unionization, since they would be unable to retain the absolute power to hire and fire in any event. Yet, by the same token, it is possible that workers would be less inclined to pay the costs of union membership if they received some purported benefit by the force of law without unionization. The ultimate weight of these considerations is an empirical question to which no easy answers appear.[49] What is clear, however, is that even if one could show that the shift in the rule either benefits or hurts unions and

---

[48] That this may be irrational does not imply that it is impossible. *See* Stigler, *Wealth, and Possibly Liberty*, 7 J. LEGAL STUD. 213 (1978).

[49] It has been reported that union leaders do not favor these reforms. Chicago Sun-Times, June 10, 1984, at 32, col. 1 (2-star ed.).

their members, the answer would not justify the rule, for it would not explain why the legal system should try to skew the balance one way or the other. The bottom line therefore remains unchanged. The case for a legal requirement that renders employment contracts terminable only for cause is as weak after distributional considerations are taken into account as before.

IV. Exceptions to the Contract at Will

An examination of the contracting objectives of parties explains why contracts at will are common. The same set of considerations, however, also helps explain why contracts at will are *not* found in all employment contexts, but are instead sometimes displaced by more elaborate contractual mechanisms. The central point is that the contract at will works only where performance on both sides takes place in lockstep progression. This condition will be satisfied where neither side has performed or where the worker's past performance has been matched by appropriate payment from the employer. In these cases the contract at will provides both employer and employee with a simple, informal "bond" against the future misfeasance of the other side: fire or quit. Where the sequence of performance requires one side to perform in full before the other side begins performance, this bonding mechanism will break down because there are no longer two unperformed promises of roughly equal value to stand as security for each other. That is why an employee will have to resort to legal action if the employer simply refuses to pay wages for work that has already been done. It is also why a contract at will cannot handle the question of compensation for job-related personal injuries, for after injury the value of the right to quit no longer balances off the right to fire.[50]

The same problem of imperfect bonding under the contract at will also arises where the nature of the employment requires work for extended periods of time. Thus, in the traditional apprenticeship contracts explicit provision had to be made to ensure that the indentured servant would not go elsewhere and likewise that the master would not abuse him during the period of service.[51] Simi-

---

[50] For an account of the elaborate consensual arrangements that grew up to handle this problem, see Epstein, *The Historical Origins and Economic Structure of Workers' Compensation Law*, 16 Ga. L. Rev. 775, 789-94, 798-803 (1982).

[51] The power of the common law's presumption in favor of at-will contracts is shown, however, by the courts' refusal to infer such terms. *See* 6 C.B. Labatt, Commentaries on the Law of Master and Servant § 2198 (2d ed. 1913). At various times, legislatures regu-

larly in earlier times, contracts for agricultural workers were understood to be for the year or for the growing season.[52] These arrangements accordingly contained extensive risk-sharing features. The worker received some interim compensation, typically in the form of room and board, that functioned as progress payments for services already rendered. Yet the contract often called for the payment of a large cash sum at the end of the harvest because the power of the employer to withhold some of the wages until that time was necessary to counteract the possibility that a worker, who had pocketed wages in advance, might abandon the employer in order to obtain high wages for day-work at harvest time when labor was at a premium.[53]

The same problems can exist with modern employment contracts. Suppose that a worker has put in the effort to obtain for the firm a large contract on which he is to be paid a commission. If the firm dismisses him under an at-will contract before the sale is consummated and the commission is formally due, most courts will (rightly) imply a term of good faith that gives the employee the commission for the work done, unless the agreement explicitly provides otherwise. Thus in *Coleman v. Graybar Electric Co.*,[54] the plaintiff's claim for compensation rested in part upon commissions that were paid annually based upon the sales record in the previous period. The court construed the contract to preclude the at-will norm: "We conclude that in this case the contract did not authorize the forfeiture of additional compensation provided in the plan of compensation if the services of the employee were terminated arbitrarily and without just cause."[55]

As the size of possible commissions increases, moreover, the

---

lated the duration of indentures. *See id.* § 2113.

[52] *See* Feinman, *supra* note 4, at 120. There is, however, no need to presume exploitation to account for this form of contract. *See infra* text following note 52.

[53] Usually the courts refused to allow the worker to recover on a quantum meruit count for the value of the services rendered when he quit before term. *See, e.g.*, Britton v. Turner, 6 N.H. 481, 486 (1834) (allowing the action but recognizing its deviation from the clear weight of authority); *see also* F. Kessler & G. Gilmore, Contracts, Cases and Materials 878 (2d ed. 1970). The majority result tends to be sound even though it requires the worker to continue the work in order to claim the deferred benefits and thus could result in some windfall to the employer. Where the worker quits before the end of the term, he can often obtain higher short-term wages in the market to offset the loss of the final payment. A rule that awarded some fraction of the deferred payment could give the worker both, which would provide him with the wrong incentives. Moreover, even when the employee loses on the wage claim, he has still obtained board and lodging, the value of which the employer would never recover by suit against the worker in breach, be it for legal or practical reasons.

[54] 195 F.2d 374 (5th Cir. 1952).

[55] *Id.* at 378.

potential gains from post-contractual exploitation will increase as well. But it does not follow that the parties are helpless to protect themselves against exploitation. Rather, when the stakes are high, it becomes worthwhile for the parties to fashion explicit allocations of the commission in the event of an employee dismissal.[56] Thus in *Fortune v. National Cash Register Co.*,[57] the commission contract in question was for the sale of a large cash register system. Under this contract, a salesman received seventy-five percent of the commission derived from a sale if the sales territory had been assigned to him at the date of the order; he received twenty-five percent of the commission if the territory had been assigned to him at the date of delivery; and he got the full commission if the territory had been assigned to him at both times.[58] The plaintiff in the action was a salesman who had been fired after the original deal had been struck but before it had been completed.[59] The seventy-five percent commission earned under the agreement had been paid, while the remaining twenty-five percent commission was paid by the firm to another employee, according to the express terms of the contract.[60] The court held that these express provisions were not dispositive and read into the agreement a covenant of good faith and fair dealing, on which it held that the plaintiff was entitled to a jury trial.[61]

The decision seems wrong in principle. The contractual provisions concerning commissions represent a rough effort to match payment with performance where the labor of more than one individual was necessary to close the sale. The case is not simply one where a strategically timed firing allowed the company to deprive a dismissed employee of the benefits due him upon completion of performance. Indeed, the firm kept none of the commission at all, so that when the case went to the jury, the only issue was whether the company should be called upon to pay the same commission twice. The court in *Fortune* did not try to understand the commis-

---

[56] The same problem commonly arises in brokerage cases where the owner tries to dismiss the broker after the buyer is located but before the agreement is concluded. A covenant of good faith is normally appropriate here to prevent the expropriation of labor by the owner-seller, and these provisions are now commonplace in brokerage agreements, which also typically provide that the brokers obtain no commission for buyers whom the seller independently locates after the brokerage period. For a collection of cases, see F. KESSLER & G. GILMORE, *supra* note 53, at 337.

[57] 373 Mass. 96, 364 N.E.2d 1251 (1977).

[58] *Id.* at 97-98, 364 N.E.2d at 1253.

[59] *Id.* at 100, 364 N.E.2d at 1254.

[60] *Id.* at 99, 364 N.E.2d at 1254.

[61] *Id.* at 101-04, 364 N.E.2d at 1255-57.

sion structure that it was prepared to condemn; instead, it made the chronic mistake of thinking that what it intuited to be an unfortunate business outcome invalidated the entire contractual structure. In its enthusiastic meddling in private contracts, the court nowhere suggested an alternative commission structure that would have better served the joint interests of the parties at the time of contract formation. Here, as in so many cases, an unquestioning adherence to the principle of freedom of contract would have yielded results both simpler and superior to those generated after an extensive but flawed judicial examination of the basic terms.

Conclusion

The recent trend toward expanding the legal remedies for wrongful discharge has been greeted with wide approval in judicial, academic, and popular circles. In this paper, I have argued that the modern trend rests in large measure upon a misunderstanding of the contractual processes and the ends served by the contract at will. No system of regulation can hope to match the benefits that the contract at will affords in employment relations. The flexibility afforded by the contract at will permits the ceaseless marginal adjustments that are necessary in any ongoing productive activity conducted, as all activities are, in conditions of technological and business change. The strength of the contract at will should not be judged by the occasional cases in which it is said to produce unfortunate results, but rather by the vast run of cases where it provides a sensible private response to the many and varied problems in labor contracting. All too often the case for a wrongful discharge doctrine rests upon the identification of possible employer abuses, as if they were all that mattered. But the proper goal is to find the set of comprehensive arrangements that will minimize the frequency and severity of abuses by employers and employees alike. Any effort to drive employer abuses to zero can only increase the difficulties inherent in the employment relation. Here, a full analysis of the relevant costs and benefits shows why the constant minor imperfections of the market, far from being a reason to oust private agreements, offer the most powerful reason for respecting them. The doctrine of wrongful discharge is the problem and not the solution. This is one of the many situations in which courts and legislatures should leave well enough alone.

VOLUME 100 JUNE 1987 NUMBER 8

HARVARD LAW REVIEW

ARTICLE

MARKET-INALIENABILITY

*Margaret Jane Radin**

Things that may be given away but not sold are market-inalienable. In this Article, Professor Radin explores the significance of market-inalienability and its justifications. The author considers and rejects two archetypes that fail to recognize market-inalienability as a separate category of social interaction. One, universal commodification, holds that everything should in principle be subject to market transfer; the other, universal noncommodification, holds that the market should be abolished. Professor Radin also explores and ultimately rejects attempts based on economic analysis and liberal philosophical doctrines to justify particular distinctions between things that are and things that are not appropriately traded in markets. She then offers an alternative justification for market-inalienability that relates it to an ideal of human flourishing. This theory takes into account both the rhetoric in which we conceive of ourselves and our situation in nonideal circumstances. Professor Radin concludes by demonstrating how the theory might be applied to three contested market-inalienabilities: prostitution, baby-selling, and surrogate motherhood.

SINCE the declaration of "unalienable rights" of persons at the founding of our republic,[1] inalienability has had a central place in our legal and moral culture. Yet there is no one sharp meaning for the term "inalienable." Sometimes inalienable means nontransferable;[2]

---

* Professor of Law, University of Southern California Law Center. I gratefully acknowledge the support of the University of Southern California Faculty Research and Innovation Fund in the preparation of this Article. Earlier versions were presented to workshops at the University of Wisconsin School of Law, Northwestern School of Law, and the University of Southern California Law Center, as well as to the Los Angeles Feminist Legal Scholars and to my Spring 1987 seminar in property theory. The Article benefited greatly from the responses of the participants. It also benefited greatly from the willingness of friends and colleagues — too numerous to name — to think and argue with me, sharing generously their time and talents. I am grateful to all of them and hope they will take up where I leave off. For making this work possible, I record my thanks to my family: Layne Leslie Britton, Wayland Jeremiah Radin, and Amadea Kendra Britton.

[1] The Declaration of Independence para. 2 (U.S. 1776).

[2] *See, e.g.*, McConnell, *The Nature and Basis of Inalienable Rights*, 3 LAW & PHIL. 25, 27 (1984) ("That which is inalienable . . . is not transferable to the ownership of another.").

sometimes only nonsalable.[3] Sometimes inalienable means nonrelinquishable by a rightholder;[4] sometimes it refers to rights that cannot be lost at all.[5] In this Article I explore nonsalability, a species of inalienability I call market-inalienability. Something that is market inalienable is not to be sold, which in our economic system means it is not to be traded in the market.

Controversy over what may be bought and sold — for example, blood or babies — pervades our news. Although some scholars have considered whether such things may be traded in markets, they have not focused on the phenomenon of market-inalienability. About fifteen years ago, for example, Richard Titmuss advocated in his book, *The Gift Relationship*,[6] that human blood should not be allocated through the market; others disagreed.[7] More recently, Elisabeth Landes and Richard Posner suggested the possibility of a thriving market in infants,[8] yet most people continue to believe that infants should not be allocated through the market.[9] What I believe is lacking, and wish to supply, is a general theory that can illuminate these debates. Two

---

[3] *See, e.g.*, Calabresi & Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972) ("An entitlement is inalienable to the extent that its transfer is not permitted between a willing buyer and a willing seller."). For discussion of Calabresi and Melamed's view of inalienability, see pp. 1864–66 below.

[4] *See, e.g.*, Barnett, *Contract Remedies and Inalienable Rights*, 4 SOC. PHIL. & POL'Y 179, 185 (1986) ("To characterize a right as inalienable is to claim that the consent of the right-holder is insufficient to extinguish the right or to transfer it to another."); Kuflik, *The Utilitarian Logic of Inalienable Rights*, 97 ETHICS 75, 75 (1986) ("An inalienable right is a right that a person has no right to give up or trade away."); Meyers, *The Rationale for Inalienable Rights in Moral Systems*, 7 SOC. THEORY & PRAC. 127, 127 (1981) ("Inalienable rights are rights that cannot be relinquished by the individuals who possess them."). For further definitions of different kinds of inalienability, see pp. 1852–55 below.

[5] *See, e.g.*, D. MEYERS, INALIENABLE RIGHTS: A DEFENSE 4 (1985) ("[A]n inalienable right is one that the right-holder cannot lose regardless of what he does or or how others treat him and even if others are justified in declining to grant him what he demands in exercising his right."); Brown, *Inalienable Rights*, 64 PHIL. REV. 192, 192 (1955) ("[I]f there are any rights properly called 'inalienable,' assertions of these rights cannot, for any reason under any circumstances, be denied.").

[6] R. TITMUSS, THE GIFT RELATIONSHIP: FROM HUMAN BLOOD TO SOCIAL POLICY (1971).

[7] *See, e.g.*, Arrow, *Gifts and Exchanges*, 1 PHIL. & PUB. AFF. 343 (1972).

[8] *See* Landes & Posner, *The Economics of the Baby Shortage*, 7 J. LEGAL STUD. 323 (1978). Elsewhere Posner said:

> That there are many people who are capable of bearing children but who do not want to raise them, and many other people who cannot produce their own children but want to raise children in their homes, suggests the possibility of a thriving market in babies, especially since the costs of production by the natural parents are typically much lower than the value that many childless people attach to the possession of children.

R. POSNER, ECONOMIC ANALYSIS OF LAW 113 (2d ed. 1977). In the most recent edition of this book, the word "thriving" has been deleted from this passage, perhaps indicating that Posner has modified his views. *See* R. POSNER, *supra*, at 139 (3d ed. 1986). I shall return to Posner's views on baby-selling below. *See infra* note 51 and accompanying text. (Further references to *Economic Analysis of Law* will be to the third edition.)

[9] *See, e.g.*, Prichard, *A Market for Babies?*, 34 U. TORONTO L.J. 341, 348–57 (1984).

possibilities for filling this theoretical gap are traditional liberalism and modern economic analysis, but in this Article I shall find them both wanting.

The most familiar context of inalienability is the traditional liberal triad: the rights to life, liberty, and property. To this triad, liberalism juxtaposes the most familiar context of alienability: traditional property rights. Although the right to hold property is considered inalienable in traditional liberalism, property rights themselves are presumed fully alienable, and inalienable property rights are exceptional and problematic.

Economic analysis, growing out of the liberal tradition, tends to view all inalienabilities in the way traditional liberalism views inalienable property rights. When it does this, economic analysis holds fast to one strand of traditional liberalism, but it implicitly rejects — or at least challenges — another: the traditional distinction between inalienable and alienable kinds of rights. In conceiving of all rights as property rights that can (at least theoretically) be alienated in markets, economic analysis has (at least in principle) invited markets to fill the social universe. It has invited us to view all inalienabilities as problematic.

In seeking to develop a theory of market-inalienability, I argue that inalienabilities should not always be conceived of as anomalies, regardless of whether they attach to things traditionally thought of as property. Indeed, I try to show that the characteristic rhetoric of economic analysis is morally wrong when it is put forward as the sole discourse of human life. My general view deviates not only from the traditional conception of the divide between inalienable and alienable kinds of rights, but also from the traditional conception of alienable property. Instead of using the categories of economics or those of traditional liberalism, I think that we should evaluate inalienabilities in connection with our best current understanding of the concept of human flourishing.

To develop this theory, which will help us to decide what things ought not to be bought and sold, I must lay a rather complex groundwork. In Part I, I articulate the various meanings of inalienability and introduce the idea of commodification. In Part II, I explore an economic view that sees all things as exchangeable, first reflecting generally on the rhetoric and methodology of the market and then examining how inalienability is seen as a method of correcting market failures. In Part III, I consider a critique of the economic view that would reject markets entirely. I find this utopian vision to be flawed by a pervasive problem of transition, but suggest we take seriously its philosophical connection between rhetoric and human flourishing. In Part IV, I consider the traditional liberal divide between market and nonmarket realms and show that the philosophical commitments of the liberal view have tended to push it toward the economic view.

Finally, in Part V, I advocate a nonideal, pragmatic evaluation of market-inalienabilities based on a conception of personhood or human flourishing that differs from that of traditional liberalism or economics. In developing this analysis, I attempt to address the transition problem that plagues our pursuit of social ideals. To show how the analysis I recommend might illuminate specific issues of market-inalienability that deeply trouble us, I conclude by bringing it to bear on commodification of sexuality and reproductive capacity: prostitution, baby-selling, and surrogacy.

I. MARKET-INALIENABILITY AND NONCOMMODIFICATION

In order to focus effectively on market-inalienability and its moral and social significance, it will be helpful first to have an overview of the range of meanings of inalienability, as well as an idea of the framework connecting alienability and commodification.

A. Traditional Meanings

Theorists have seldom recognized that we have no one sharp meaning of inalienability.[10] Nevertheless, the traditional meanings of inalienability share a common core: the notion of alienation as a separation of something — an entitlement, right, or attribute[11] — from its holder. Inalienability negates the possibility of separation. Meanings proliferate because the separation that constitutes alienation can be either voluntary or involuntary, and can result in the entitlement, right, or attribute ending up in the hands of another holder, or in its simply being lost or extinguished.[12] Any particular entitlement,

---

[10] Most of the definitions quoted in notes 2–5 above ignore their rivals. Among recent commentators, Susan Rose-Ackerman is exceptional in recognizing many meanings. *See* Rose-Ackerman, *Inalienability and the Theory of Property Rights*, 85 COLUM. L. REV. 931 (1985); *infra* note 67 (discussing Rose-Ackerman's approach to the meanings of inalienability).

[11] The traditional conception of alienation as separation of objects from persons is related to the traditional subject/object dichotomy, which I discuss in Section B of Part III below. In what follows, I shall often refer generally to whatever is inalienable as a "thing." This is a necessary shorthand, although it does present the danger of an unwanted connotation of objectness, especially in light of the subject/object dichotomy.

[12] As these two variables suggest, each of four broad categories of separability might be negated by a corresponding form of inalienability: involuntary extinguishment (cancellation, forfeiture of civil rights); voluntary extinguishment (waiver, abandonment); involuntary transfer (condemnation, adverse possession); and voluntary transfer (gift, sale). The negation of involuntary transfer is less likely than the other categories to be conceived of as inalienability, for example, we do not consider the prevention of theft an inalienability. There are other variables that are sometimes significant for understanding inalienabilities. The most important of them are: the sovereign's role in the interaction (whether the sovereign is the instrument of involuntary loss or the recipient of involuntary transfer); the nature of the holder of an inalienable right, entitlement, or attribute (whether the holder is a person or group, and whether the person has an official capacity or the group has special normative significance); and the availability of

right, or attribute may be subject to one or more forms of inalienability.

In one important set of meanings, inalienability is ascribed to an entitlement, right, or attribute that cannot be lost or extinguished. If involuntary loss is its focus, inalienable may mean nonforfeitable[13] or noncancelable; if voluntary loss is its focus, inalienable may mean nonwaivable[14] or nonrelinquishable.

In another important set of meanings, inalienability is ascribed to an entitlement, right, or attribute that cannot be voluntarily transferred from one holder to another.[15] Inalienability in these uses may mean nongiveable, nonsalable, or completely nontransferable. If something is nontransferable, the holder cannot designate a successor holder.[16] Nongiveability and nonsalability are subsets of nontransferability. If something is inalienable by gift, it might be transferred by sale;[17] if it is inalienable by sale, it might be transferred by gift.[18] This nonsalability is what I refer to as market-inalienability. In precluding sales but not gifts, market-inalienability places some things outside the marketplace but not outside the realm of social intercourse.

---

compensation. These and other variables can be thought either to create a larger matrix or to delineate subcategories in the four broad categories. Market-inalienability is a normatively important subcategory of inalienabilities that negate voluntary transfer, a category delineated by the distinction between monetary exchanges and other voluntary transfers. In the text, I do not seek to elaborate on the matrix of inalienabilities, but rather to set the scene for an exploration of market-inalienability. The taxonomy of inalienabilities I propose here should be compared with that of Susan Rose-Ackerman, cited in note 10 above, at 933–35. *See infra* note 67.

[13] Nonforfeitability is ambiguous. It may refer to an entitlement, right, or attribute that cannot be involuntarily negated, such as certain civil rights; or it may refer to things that cannot be involuntarily transferred to the sovereign without compensation. In the text I use the former meaning.

[14] Waiver is ambiguous. It may refer either to permanent or temporary abrogation of an entitlement, right, or attribute. In the latter meaning, perhaps we should not speak of waiver as "loss." In the text I use the former meaning.

[15] Preclusion of involuntary transfer is not usually conceived of as inalienability. *See supra* note 12.

[16] Something that is inalienable in this sense need not be inalienable in the broader senses detailed above. *See supra* note 12. Something that is nontransferable might still be forfeited, canceled, relinquished, waived, or perhaps involuntarily transferred to the government or its designates.

[17] In addition to transfer by gift and sale, barter also represents a theoretically possible means of transfer. I do not consider barter, however, because it is not a widespread method of exchange in our culture.

[18] There are also subsets of gift transfer: transfer *inter vivos*, and bequest or devise. Although gift transmission at death is not a focus of this Article, market-inalienability may sometimes leave open both kinds of gift transfer; for example, one may will one's body or organs. In other cases, bequest might be irrelevant, as with sexual services. *See infra* pp. 1921–25. Bequest might be prohibited in cases in which it imparts unwanted connotations of property, for example, in the case of relinquishment of children for adoption; a testator can create a guardianship for a child but may not will the child itself to the guardian. *See infra* pp. 1925–28.

Market-inalienability negates a central element of traditional property rights, which are conceived of as fully alienable.[19] But market-inalienability differs from the nontransferability that characterizes many nontraditional property rights — entitlements of the regulatory and welfare state — that are both nongiveable and nonsalable.[20] Market-inalienability also differs from the inalienability of other things, like voting rights, that seem to be moral or political duties related to a community's normative life; they are subject to broader inalienabilities that preclude loss as well as transfer.[21] Unlike the inalienabilities attaching to welfare entitlements or political duties, market-inalienability does not render something inseparable from the person, but rather specifies that market trading may not be used as a social mechanism of separation. Finally, market-inalienability differs from the inalienability of things, like heroin, that are made nontransferable in order to implement a prohibition,[22] because it does not signify that something is social anathema. Indeed, preclusion of sales often coexists with encouragement of gifts. For example, the market-

---

[19] Traditional property rights are alienable in all senses except cancellation; they may be forfeited, relinquished, waived, condemned, and transferred by both gift and sale. The unavailability of cancellation is usually not by itself considered inalienability. Yet it is helpful to think of noncancelability as inalienability in order to see the scope of the concept. Indeed, the liberal "inalienable right to property" may just mean that liberal alienable property rights are noncancelable. (It may mean other things as well, for example, that the right of the autonomous individual to be a property holder is both nonrelinquishable and noncancelable. *See infra* p. 1900.

[20] Examples are entitlements to social security and welfare benefits, and many kinds of licenses. I think of these as status-inalienabilities because they rigidify possession, constraining or precluding change, signifying some strong form of inseparability from the holder. Forms of status-inalienability could range from prohibition of voluntary transfers among private parties to prohibition of any kind of loss.

[21] Nontransferable rights that at the same time may implicate affirmative duties fall into a category I think of as community-inalienability. Examples are the right-duty to vote in political elections and the right-duty to become educated. Rights of this kind not only may not be lost through change of hands, extinguishment, or cancellation, but also ought to be exercised. Although community-inalienability is a convenient label for these rights that are simultaneously duties, the more communitarian one's views about the nature of the person and the nature of social life, the more all justifiable inalienabilities will be related to community. For example, Laurence Tribe argues that "rights that are relational and systemic are *necessarily* inalienable: individuals cannot waive them because individuals are not their sole focus." Tribe, *The Abortion Funding Conundrum: Inalienable Rights, Affirmative Duties, and the Dilemma of Dependence*, 99 HARV. L. REV. 330, 333 (1985) (applying this reasoning to the right to choose abortion) (emphasis in original).

[22] Some things are deemed socially unacceptable to possess, give, or sell; their existence is denounced completely by the social order. Heroin is in this class; alcoholic beverages passed into and out of it. The inalienability of things in this class is subsidiary to a social attempt to obliterate them. It is illegal to sell heroin only because we want no one to have anything to do with heroin. Growing heroin, possessing heroin, and giving away heroin are prohibited too. To distinguish inalienabilities incident to prohibitions from other kinds of inalienabilities, the former can be labeled prohibition-inalienabilities.

inalienability of human organs does not preclude — and, indeed, may seek to foster — transfer from one individual to another by gift.[23]

B. The Commodification Issue

Market-inalienability often expresses an aspiration for noncommodification. By making something nonsalable we proclaim that it should not be conceived of or treated as a commodity.[24] When something is noncommodifiable, market trading is a disallowed form of social organization and allocation. We place that thing beyond supply and demand pricing, brokerage and arbitrage, advertising and marketing, stockpiling, speculation, and valuation in terms of the opportunity cost of production.

Market-inalienability poses for us more than the binary choice of whether something should be wholly inside or outside the market, completely commodified or completely noncommodified. Some things are completely commodified — deemed suitable for trade in a laissez-faire market.[25] Others are completely noncommodified — removed from the market altogether. But many things can be described as incompletely commodified — neither fully commodified nor fully removed from the market.[26] Thus, we may decide that some things should be market-inalienable only to a degree, or only in some aspects.

---

[23] *See* National Organ Transplant Act of 1984, 42 U.S.C. § 274(e) (1982) (banning organ sales in interstate commerce). Whether organ sales are morally permissible or should legally be permitted is currently controversial. *See* Andrews, *My Body, My Property*, 16 HASTINGS CENTER REP., Oct. 1986, at 28, 36 (arguing thoughtfully for a "quasi-property" approach in which "human beings have the right to treat certain physical parts of their bodies as objects for possession, gift, and trade"). In my terms Andrews's position is actually an incomplete commodification, *see* pp. 1917–21, because it precludes brokering of organs and treatment of one's body parts as property by other people. *Cf.* Murray, *The Gift of Life Must Always Remain a Gift*, 7 DISCOVER, Mar. 1986, at 90; Comment, *Retailing Human Organs Under the Uniform Commercial Code*, 16 J. MARSHALL L. REV. 393, 405 (1983) (arguing that "society should not view the sale of human organs any differently than the sale of other necessary commodities such as food, shelter, and medication"). This debate is an example of contested commodification. *See infra* pp. 1856–58.

[24] As I use it here, the term "commodity" means simply something that is thought appropriate to buy and sell through a market. Later I discuss further complexities of meaning in the term "commodification." *See infra* pp. 1859–60. For now, note that this definition makes it awkward to speak of nonsalable or nonmarket commodities. I think it appropriate to restrict the term "commodity" to monetary trade and its rhetoric, so that when speakers do apply the term to nonmarket activities, we can identify its use as market rhetoric. If someone says, "Love is a scarce commodity," this definition will enable us to see clearly that she is speaking about love as if it were a resource available on the market. Market rhetoric is discussed in Section B of Part III below.

[25] *See infra* p. 1861.

[26] For a discussion of incomplete commodification, see pp. 1917–21 below. Things that are incompletely commodified do not fully exhibit the typical indicia of traditional property and contract. For example, things that are subject to price controls are incompletely commodified because freedom to set prices is part of the traditional understanding of property and contract.

To appreciate the need to develop a satisfactory analysis of market-inalienability, consider the deeply contested issues of commodification that confront us. Infants and children,[27] fetal gestational services,[28] blood,[29] human organs,[30] sexual services,[31] and services of college athletes[32] are some salient things whose commodification is contested.[33] Our division over whether to place a monetary equivalent on a spouse's professional degree[34] or homemaker services in a divorce;

See, e.g., Block v. Hirsh, 256 U.S. 135, 159 (1921) (McKenna, J., dissenting) (protesting that rent control "is contrary to every conception of leases that the world has ever entertained").

[27] *Compare* Landes & Posner, *supra* note 8 (suggesting a free market in babies as an experimental solution to the current baby shortage) *with* Prichard, *supra* note 9 (outlining objections to a free market in babies based on market failure, degradation, and a child-centered view of adoption mechanisms). Baby-selling is discussed at pp. 1925–28 below.

[28] *Compare* P. SINGER & D. WELLS, MAKING BABIES: THE NEW SCIENCE AND ETHICS OF CONCEPTION (1985) (arguing in favor of heavily regulated surrogacy arrangements) *with* Krimmel, *The Case Against Surrogate Parenting*, 13 HASTINGS CENTER REP., Oct. 1983, at 35 (objecting to the legalization of surrogacy arrangements because they may cause social ills related to eugenics, family breakdown, and psychic harm to children). Surrogacy is discussed at pp. 1928–36 below.

[29] *Compare* R. TITMUSS, *supra* note 6 (arguing that blood should not be salable because sales repress altruism and erode feelings of community), *and* Singer, *Freedom and Utilities in the Distribution of Health Care*, in MARKETS AND MORALS 149 (G. Bermant, P. Brown & G. Dworkin eds. 1977) (maintaining that a system of voluntary donations of blood is more efficient than a system of sales, and additionally that it fosters feelings of community) *with* Arrow, *supra* note 7 (criticizing Titmuss's conclusions). The argument that blood should be market-inalienable in order to preserve opportunities for altruism is discussed at pp. 1913–14 below.

[30] *See* sources cited *supra* note 23.

[31] *Compare* D. RICHARDS, SEX, DRUGS, DEATH, AND THE LAW 84–127 (1982) (arguing that Kantian autonomy, rightly understood, would permit commercial sex), *and* Ericsson, *Charges against Prostitution: An Attempt at a Philosophical Assessment*, 90 ETHICS 335 (1980) (arguing that a legalized market would cure many of the evils currently associated with prostitution) *with* Pateman, *Defending Prostitution: Charges Against Ericsson*, 93 ETHICS 561 (1983) (rejecting Ericsson's argument as based on a misunderstanding of the feminist critique of prostitution). Sale of sexual services is discussed at pp. 1921–25 below.

[32] *See* D. EITZEN & G. SAGE, SOCIOLOGY OF AMERICAN SPORT 53–68 (3d ed. 1986) (describing the ideal of amateurism in sports); Koch, *A Troubled Cartel: The NCAA*, 38 LAW & CONTEMP. PROBS. 135 (1973); Note, *Compensation for College Athletes: A Run for More than the Roses*, 22 SAN DIEGO L. REV. 701 (1985) (arguing that compensation of college athletes should be regarded as legitimate); *infra* note 262.

[33] In addition, there is a debate over possible sale of sperm, eggs, and embryos. *See, e.g.*, Andrews, *supra* note 23, at 33; P. SINGER & D. WELLS, *supra* note 28. There are numerous other aspects of health care, such as the allocation of artificial organs or kidney dialysis service, whose suitability for a market regime is intensely debated. The debate over the degree to which the indicia of traditional property should attach to a celebrity's "persona" is also an example of contested commodification. *See, e.g.*, Terrell & Smith, *Publicity, Liberty, and Intellectual Property: A Conceptual and Economic Analysis of the Inheritability Issue*, 34 EMORY L.J. 1 (1985). Contested commodification can be viewed historically. For example, child labor and public offices used to be bought and sold. They passed through a period of contest and were decommodified. *See* V. ZELIZER, PRICING THE PRICELESS CHILD (1985); Nelson, *Officeholding and Powerwielding: An Analysis of the Relationships Between Structure and Style in American Administrative History*, 10 LAW & SOC'Y 187 (1976).

[34] *See, e.g.*, Krauskopf, *Recompense for Financing Spouse's Education: Legal Protection for*

or on various kinds of injuries in tort actions, such as loss of consortium, is another form of contest over commodification.[35] Monetization — commodification — of clean air and water is likewise deeply contested.[36] Moreover, debates about some kinds of regulation can be seen as contested incomplete commodification, with the contest being over whether to allow full commodification (a laissez-faire market regime) or something less. If we see the debates this way, residential rent control, minimum wage requirements, and other forms of price regulation, as well as residential habitability requirements, safety regulation, and other forms of product-quality regulation all become contests over the issue of commodification.[37]

How are we to determine the extent to which something ought to be noncommodified, so that we can determine to what extent market-inalienability is justified? Because the question asks about the appropriate relationship of particular things to the market, normative theories about the appropriate social role of the market should be helpful in trying to answer it. We can think of such theories as ordered on a continuum stretching from universal noncommodification (nothing in markets) to universal commodification (everything in markets). On this continuum, Karl Marx's theory can symbolize the theoretical pole of universal noncommodification, and Richard Posner's can be seen as close to the opposite theoretical pole.[38] Distributed along the con-

---

the Marital Investor in Human Capital, 28 U. KAN. L. REV. 379 (1980). *Compare* O'Brien v. O'Brien, 66 N.Y.2d 576, 489 N.E.2d 712 (1985) (holding that a medical license acquired during marriage is marital property subject to equitable distribution upon divorce) *with In re* Graham, 194 Colo. 429, 574 P.2d 75 (1978) (holding that an M.B.A. degree acquired during marriage is not an item of property and cannot be distributed at divorce).

[35] The debate over commodification in tort law is touched upon below. *See infra* pp. 1876–77.

[36] *See, e.g.*, S. KELMAN, WHAT PRICE INCENTIVES?: ECONOMISTS AND THE ENVIRONMENT 27–91 (1981).

[37] For a discussion of this view of regulation, see pp. 1918–21 below. Prohibition debates can also be cast in the form of contested commodification, if one sees the market as encompassing the whole of social life. The Supreme Court once saw the existence of pornography as lowering the "tone of commerce," perhaps equating commerce with markets and markets with the social arena. *See* Paris Adult Theatre I v. Slaton, 413 U.S. 49, 58 (1973).

[38] Posner's tendency toward universal commodification can be seen most readily in his "universality" and "transferability" criteria for an appropriate property regime. *See* R. POSNER, *supra* note 8, at 29–33. His discussion of an economic theory of property

> implies that if every valuable (meaning scarce as well as desired) resource were owned by someone (universality), ownership connoted the unqualified power to exclude everybody else from using the resource (exclusivity) as well as to use it oneself, and ownership rights were freely transferable, or as lawyers say alienable (transferability), value would be maximized.

Id. at 32. The only limitation Posner places on this claim that everything valuable should be alienable property is that it must be qualified by the costs of implementing such a system. *See id.*

Posner's tendency toward universal commodification can also be seen in his definition of "value" in terms of money. *See id.* at 11; *cf.* R. POSNER, THE ECONOMICS OF JUSTICE 115

tinuum are theorists we may call pluralists — those who see a normatively appropriate but limited realm for commodification coexisting with one or more nonmarket realms. Pluralists often see one other normative realm besides that of the market, and partition the social world into markets and politics,[39] markets and rights,[40] or markets and families;[41] but pluralists also may envision multiple nonmarket realms.[42] For a pluralist, the crucial question is how to conceive of the permissible scope of the market. An acceptable answer would solve problems of contested commodification.

(1981) (defending wealth-maximization as "the criterion for judging whether acts and institutions are just or good"). He defines economics globally as "the science of rational choice" in a world of scarce resources, its task being "to explore the implications of assuming that man is a rational maximizer of his ends in life, his satisfactions — what we shall call his 'self-interest.'" R. POSNER, *supra* note 8, at 3; *cf.* R. POSNER, THE ECONOMICS OF JUSTICE, *supra*, at 1–5 (defending the application of economics to all fields of human activity).

Although it seems that Posner is as close to the universal commodification pole of the hypothetical continuum as any theorist, he occasionally states that economic analysis may have some limits. *See, e.g.*, R. POSNER, *supra* note 8, at 25–26. Even when he admits limits, however, he seems to deny them again, as in this passage:

> There may well be definite although wide boundaries on both the explanative and reformative power of economic analysis of law. Always, however, economics can provide value clarification by showing the society what it must give up to achieve a noneconomic ideal of justice. The demand for justice is not independent of its price.

Id. at 26; *cf. id.* at 244 (referring to "corrective justice" as a "commodity").

The writings of Gary Becker also exemplify the universal commodification pole. *See, e.g.*, G. BECKER, A TREATISE ON THE FAMILY (1981) (developing an economic approach to marriage, family, and procreation). For further discussion of universal commodification, see Part II below.

Although a typical ideal of universal commodification is, as Posner states in the first passage quoted, that everything scarce that people value should be ownable and salable, Frank Michelman argues that universal entitlement and free alienation are theoretically impossible. *See* Michelman, *Ethics, Economics and the Law of Property*, in ETHICS, ECONOMICS AND THE LAW: NOMOS XXIV 3 (J. Pennock & J. Chapman eds. 1982). Gregory Alexander has shown the tension in nineteenth-century property law caused by this theoretical problem. *See* Alexander, *The Dead Hand and the Law of Trusts in the Nineteenth Century*, 37 STAN. L. REV. 1189 (1985).

[39] *See, e.g.*, C. LINDBLOM, POLITICS AND MARKETS (1977). Those who speak of *Gemeinschaft* and *Gesellschaft* are also pluralists if these two forms of interaction are conceived of as properly coexisting. *See* F. TONNIES, COMMUNITY AND SOCIETY (C. Loomis trans. & ed. 1963).

[40] *See, e.g.*, A. OKUN, EQUALITY AND EFFICIENCY: THE BIG TRADEOFF (1975).

[41] *See, e.g.*, Olsen, *The Family and the Market: A Study of Ideology and Legal Reform*, 96 HARV. L. REV. 1497 (1983) (describing and criticizing the prevalent ideology in which the family and the market form a dichotomy, related to ideological dichotomies between the state and civil society and between male and female).

[42] Michael Walzer, one of the more interesting theorists who view society in this compartmentalized way, distinguishes 11 separate spheres of justice: membership (citizenship), security and welfare, money and commodities (the market), office, hard work (distasteful or dangerous but socially necessary tasks), free time, education, kinship and love, divine grace (religious freedom), recognition (equal respect), and political power. *See* M. WALZER, SPHERES OF JUSTICE (1983). In my view, Walzer's theory is flawed because it assumes that a free-market sphere is presumptively just. *See* Radin, *Justice and the Market Domain*, in MARKETS AND JUSTICE: NOMOS XXXI (J. Chapman ed. forthcoming).

Pluralism with its crucial question is a main focus of this Article, because a species of pluralism has been prevalent in liberal thought, and because pluralism is a common-sense position for many people. In order to explore pluralism, both in its traditional form and as it might be reconceived to yield acceptable answers, it will first be necessary to review a modern alternative to pluralism — universal commodification in the form of economic analysis — and the critique of this alternative.

II. UNIVERSAL COMMODIFICATION

Under universal commodification, there is no deep question about the appropriate scope of the market, because the market is theoretically all-encompassing. From this point of view, all inalienabilities reduce to market-inalienability, and market theory itself, using a market failure analysis, can determine when things should not be bought and sold.

A. The Rhetoric and Methodology of the Market

The term "commodification" can be construed narrowly or broadly. Narrowly construed, commodification describes actual buying and selling (or legally permitted buying and selling) of something.[43] Broadly construed, commodification includes not only actual buying and selling, but also market rhetoric, the practice of thinking about interactions as if they were sale transactions, and market methodology, the use of monetary cost-benefit analysis to judge these interactions. Universal commodification embraces this broad construction in its most expansive form, limiting actual buying and selling only by the dictates of market methodology, and solving problems of contested commodification by making everything in principle a commodity.[44]

---

[43] There are actually three gradations of commodification in the narrow sense: (1) sales that are legally or morally permitted; (2) sales that take place in spite of illegality or immorality (black markets); and (3) "sales" caused by official monetization of nonmonetary interests (for example, compensation in tort). In this Article I am more interested in the first and third aspects of this narrow view of commodification, because I am concerned with how we might evaluate what things ought not to be commodified, even if some people do violate the strictures. Yet, wholesale violation of the strictures will not be irrelevant to nonideal evaluation. *See infra* p. 1921.

[44] A commodity, at least as I am using the term here, has the following indicia pertaining to individual and social value. From the social point of view, the value of a commodity is defined as its exchange value, often referred to as market value, when it is traded in a laissez-faire market or hypothetically traded in a hypothetical laissez-faire market. Exchange value is expressed in money. Hence, under universal commodification, all social value is capable of being expressed in money terms. Moreover, all commodities are fungible — capable of being

Universal commodification means that anything some people are willing to sell and others are willing to buy in principle can and should be the subject of free market exchange. Moreover, universal commodification means that everything people need or desire, either individually or in groups, is conceived of as a commodity. "Everything" includes not only those things usually considered goods, but also personal attributes, relationships, and states of affairs. Under universal commodification, the functions of government, wisdom, a healthful environment, and the right to bear children are all commodities.[45]

---

reduced to money without changing in value, and completely interchangeable with each and every other commodity in terms of exchange value.

From the individual point of view, the value of a commodity is defined as either the sum of money the holder will accept in order to relinquish it or the sum of money the potential holder will pay in order to acquire it. *See, e.g.*, R. POSNER, *supra* note 8, at 11. Universal commodification often tends to presume that individual value is equivalent to exchange value. When a possible divergence is recognized, exchange value is often called "objective" value and individual value is often called "subjective" value. Even if it recognizes a difference between "objective" and "subjective" value, universal commodification tends to presume that the two measures of "subjective" value are equivalent. The possible divergence between what an entitled holder would demand to relinquish something and what an unentitled potential holder would pay to acquire it is sometimes called by critics "the offer-asking problem." *See* Baker, *The Ideology of the Economic Analysis of Law*, 5 PHIL. & PUB. AFF. 3, 32–41 (1975); Kennedy, *Cost-Benefit Analysis of Entitlement Problems: A Critique*, 33 STAN. L. REV. 387, 401–21 (1981).

The holder of a commodity — that is, the person viewed as commodity holder — is defined as being indifferent among holding that particular commodity, some other commodity of equivalent value to her (in money), or the sum of money itself. Hence, under universal commodification, all things of value to the person — including personal attributes, relationships, and philosophical commitments — are described in monetary terms and are in principle alienable.

My characterization of universal commodification may be compared with Mark Kelman's description of the "core premises" of legal economics. *See* Kelman, *Misunderstanding Social Life: A Critique of the Core Premises of "Law and Economics"*, 33 J. LEGAL EDUC. 274 (1983); *see also* Harrison, *Egoism, Altruism and Market Illusions: The Limits of Law and Economics*, 33 UCLA L. REV. 1309 (1986) (arguing against the egoism and narrow self-interest assumptions of economics); Michelman, *Norms and Normativity in the Economic Theory of Law*, 62 MINN. L. REV. 1015, 1039–48 (1978) (arguing against economic reductionism and a reductionist interpretation of the role of courts); Tribe, *Ways Not to Think About Plastic Trees: New Foundations for Environmental Law*, 83 YALE L.J. 1315 (1974) (arguing against the rhetoric of human self-interest); *cf.* A. BUCHANAN, ETHICS, EFFICIENCY, AND THE MARKET (1985) (discussing varieties of moral arguments for and against the market).

[45] *See, e.g.*, A. ALCHIAN & W. ALLEN, EXCHANGE AND PRODUCTION: COMPETITION, COORDINATION, AND CONTROL 17 n.1 (3d ed. 1983) (defining "economic goods" to include "*all* things that we would like to have — friendships, cleanliness, health, honesty and the like — and not merely marketable things like milk, shoes, and cars"); *cf.* R. POSNER, *supra* note 8, at 26, 244 (quoted in note 38 above) (speaking of justice as a commodity); Johnsen, *Wealth Is Value*, 15 J. LEGAL STUD. 263, 269 n.22 (1987) (speaking of justice as a scarce good and an element of wealth). *See generally* ECONOMIC IMPERIALISM: THE ECONOMIC APPROACH APPLIED OUTSIDE THE FIELD OF ECONOMICS (G. Radnitzky & P. Bernholz eds. 1987); Hirshleifer, *The Expanding Domain of Economics*, 75 AM. ECON. REV. 53, 53 ("Special Issue" Dec. 1985) (arguing that "economics really does constitute the universal grammar of social science," because

Universal commodification is characterized by universal market rhetoric and universal market methodology. In universal market rhetoric — the discourse of complete commodification — everything that is desired or valued is conceived of and spoken of as a "good." Everything that is desired or valued is an object that can be possessed, that can be thought of as equivalent to a sum of money, and that can be alienated. The person is conceived of and spoken of as the possessor and trader of these goods, and hence all human interactions are sales.

Market methodology includes a cost-benefit analysis, evaluating human actions and social outcomes in terms of actual or hypothetical gains from trade measured in money. Under universal commodification, market trading and its outcomes represent individual freedom and the ideal for individuals and society. Unrestricted choice about what goods to trade represents individual freedom, and maximizing individual gains from trade represents the individual's ideal. All social and political interactions are conceived of as exchanges for monetizable gains. Politics reduces to "rent seeking" by log-rolling selfish individuals or groups,[46] and the social ideal reduces to efficiency.[47]

In seeking efficiency through market methodology, universal commodification posits the laissez-faire market as the rule. Laissez-faire is presumptively efficient because, under universal commodification, voluntary transfers are presumed to maximize gains from trade, and all human interactions are characterizable as trades. Laissez-faire also presumptively expresses freedom, because freedom is defined as free choices of the person seen as trader.[48]

---

its analytical categories of "scarcity, cost, preferences, opportunities . . . are truly universal in applicability"; but economists should become aware of how constraining has been "their tunnel vision about the nature of man and social interactions").

[46] *See* J. BUCHANAN & G. TULLOCK, THE CALCULUS OF CONSENT (1974); D. MUELLER, PUBLIC CHOICE (1979); Buchanan, *Rent Seeking and Profit Seeking*, in TOWARD A THEORY OF THE RENT-SEEKING SOCIETY 4 (J. Buchanan, R. Tollison & G. Tullock eds. 1980) ("Rent seeking is designed to describe behavior in institutional settings where individual efforts to maximize value generate social waste rather than social surplus."); Michelman, *Political Markets and Community Self-Determination: Competing Judicial Models of Local Government Legitimacy*, 53 IND. L.J. 145, 148 (1977–78) (describing the "public choice" model in which "the legislature is conceived as a market-like arena"); Sunstein, *Interest Groups in American Public Law*, 38 STAN. L. REV. 29 (1985).

[47] Proponents of law and economics often note that they do not endorse the view that efficiency equals justice, because an efficient state (however efficiency is defined) is always efficient relative to an initial wealth distribution, and the initial distribution may be unjust. *See, e.g.*, R. POSNER, *supra* note 8, at 13. But many of them ignore their caveat. *See, e.g.*, *id.* at 25 (stating that efficiency is "perhaps the most common" meaning of "justice").

[48] The conception of the person under universal commodification is discussed at p. 1885 below. The presumptive efficiency and presumptive freedom of laissez-faire suggest that the philosophical premises of theorists whose views tend toward universal commodification may be either utilitarian or libertarian. Many law-and-economics theorists are utilitarians. *See, e.g.*, Ellickson, *Adverse Possession and Perpetuities Law: Two Dents in the Libertarian Model of*

Universal commodification is an archetype, a caricature. Economic analysts do not explicitly embrace it, but some of them, some of the time, implicitly come close. Posner, for example, suggests that everything ought to be ownable and salable,[49] and he often seems to embrace universal market rhetoric and universal market methodology. Posner speaks in market rhetoric when he says that "the prohibition against rape is to the marriage and sex 'market' as the prohibition

---

Property Rights, 64 WASH. U.L.Q. 723, 737 (1986) (finding, with approval, that "the deep structure of property law has traditionally been . . . transaction-cost utilitarianism"); Ellickson, Remarks in *Time, Property Rights, and the Common Law* (Round Table Discussion), 64 WASH. U.L.Q. 793, 796 (1986) (suspecting that "most of us" law-and-economics scholars are utilitarians at bottom). Posner describes welfare economics as often equated with utilitarianism, but he attempts to dissociate himself at least from the classical version of utilitarianism by embracing "wealth" rather than "utility" as the ethical maximand. *See* R. POSNER, THE ECONOMICS OF JUSTICE, *supra* note 38, at 44–88. On the relationship of economics to utilitarianism, see I.M.D. LITTLE, A CRITIQUE OF WELFARE ECONOMICS 6–14 (2d ed. 1957), and Coleman, *Economics and the Law: A Critical Review of the Foundations of the Economic Approach to Law*, 94 ETHICS 649 (1984).

Some theorists whose views tend toward universal commodification see themselves as libertarians, although if pressed, the ethic that drives their analysis seems to be wealth or welfare maximization. *See, e.g.*, J. BUCHANAN, THE LIMITS OF LIBERTY (1975); R. EPSTEIN, TAKINGS: PRIVATE PROPERTY AND THE POWER OF EMINENT DOMAIN 331–50 (1985). Richard Epstein seems to have undergone an odyssey from libertarianism to utilitarianism, passing through a stage in which he tried to embrace both at once. *See infra* note 66. Even a purer kind of libertarian like Robert Nozick tends toward commodification, although Nozickian libertarianism lacks the pervasive rhetoric of monetary cost-benefit analysis and rejects welfare economics. Nozick's first premise is that people have Lockean rights, but for him justice seems to require that these rights all be voluntarily exchangeable. In fact, Nozick's conception of ideal justice, consisting of a theory of just acquisition and a theory of just transfer, is simply the infrastructure of the market: private property plus free contract. *See* R. NOZICK, ANARCHY, STATE, AND UTOPIA 150–53 (1974); *cf. infra* p. 1888.

The archetype I characterize as universal commodification is different from mere consequentialism or mere utilitarianism. Consequentialism is a very broad description for the idea of identifying good and bad by results; of course it is possible to do this without making monetization or market trading central to the scheme. Although some utilitarians may be close to universal commodification, others define individual and aggregate social value as utility or welfare maximization without supposing utility to be intrinsically characterizable in money terms and without supposing interpersonal comparisons to be possible. *See generally* A. SEN, CHOICE, WELFARE AND MEASUREMENT (1982). This type of utilitarianism lacks the aspect of fungibility that characterizes universal commodification; economists who accept the possibility of judgments calculating Kaldor-Hicks efficiency are closer to universal commodification than those who do not. *See, e.g.*, R. POSNER, *supra* note 8, at 12–15. Utilitarianism without interpersonal comparisons also tends toward universal commodification, however, because it tends to conceive of the person as an empty receptacle for undifferentiated welfare to be obtained by satisfying preferences, or for interchangeable (although unmonetized) subjective "utils." *See infra* pp. 1884–85.

[49] Posner argues that, but for the costs of implementing a property system, value would be maximized if everything scarce and desired were ownable and salable; he also argues that we ought to act so as to maximize value (wealth). *See supra* note 38. Thus, everything scarce and desirable ought to be ownable and salable.

against theft is to explicit markets in goods and services."[50] Posner uses universal market methodology to suggest that a free market in infants should replace the regulated market (adoption mechanisms) we now have and the black market engendered by evasion of it.[51]

B. Inalienability as a Means of Correcting Market Failure

Universal commodification leads to a characteristic way of understanding inalienability in general and market-inalienability in particular. First, no inalienability or restraint on alienation should exist unless market methodology itself requires it. Second, if inalienability is required, it is accounted for in market terms and described in market rhetoric. These two premises combine to produce a transaction costs model of inalienability, in which inalienability is a means of controlling externalities that prevent the market from achieving an efficient result. Third, market-inalienability is not seen as a subcategory of inalienability. When one supposes, for purposes of explanation and justification, that every human interaction is a sale, then all inalienabilities collapse into nonsalability.

---

[50] Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1199 (1985). The reference to the market for sex and marriage includes scare quotes presumably only because the market is implicit rather than explicit. Posner says that the purpose of the passage in which this statement occurs is "to point out that economic analysis need not break down in the face of such apparently noneconomic phenomena as rape." *Id.* As another example of Posner's market rhetoric, consider the passages on baby-selling quoted in note 8 above and note 51 below.

[51] *See* Landes & Posner, *supra* note 8. In their article, Landes and Posner speculate on "the possibility of taking some tentative and reversible steps toward a free baby market in order to determine experimentally the social costs and benefits of using the market in this area." *Id.* at 347; *accord* R. POSNER, *supra* note 8, at 139–43. In his book, Posner states that "[t]he baby shortage would be considered an intolerable example of market failure if the commodity were telephones rather than babies." *Id.* at 139. He "examine[s] in a scientific spirit the objections to permitting the sale of babies for adoption," *id.* at 141, and finds them all to be unpersuasive. For example, he argues:

> Opponents of the market approach also argue that the rich would end up with all the babies, or at least all the good babies . . . . Such a result might of course be in the children's best interest, but it is unlikely to materialize. Because people with high incomes tend to have high opportunity costs of time, the wealthy usually have smaller families than the poor. Permitting babies to be sold would not change this situation. Moreover, the total demand for children on the part of wealthy childless couples must be very small in relation to the supply of children, even high-quality children, that would be generated in a system where there were economic incentives to produce children for purchase by childless couples.

Id. at 142. Posner goes on to say that the poor may actually do better in a free baby market than under present adoption law, because "[p]eople who might flunk the agencies' criteria on economic grounds might, in a free market with low prices, be able to adopt children, just as poor people are able to buy color television sets." *Id.* at 143. Posner has recently said, however, that he "did not advocate a free market in babies." Posner, *Mischaracterized Views*, 69 JUDICATURE 321 (1986).

The transaction costs model is developed by Guido Calabresi and A. Douglas Melamed in their treatment of "inalienability rules."[52] Even though its discussion of inalienability is limited, their article has been seminal for those who conceive of inalienability in the market mode.[53] Calabresi and Melamed divide protection of entitlements into property rules, liability rules, and inalienability. Property rules signify a scheme of free transfers between willing sellers and buyers, with no coerced transfers; liability rules signify a scheme of allowable coerced transfers at market prices set by official entities, such as courts. Calabresi and Melamed argue that property rules are prima facie efficient and therefore desirable. Liability rules are an exception to the property-rule regime, justifiable only when transaction costs of various kinds cause market failures to undermine the prima facie efficiency of property rules.[54] Both the property-rule regime and the exception to it are generated by market methodology and the pursuit of efficiency.

Calabresi and Melamed conceive of inalienability as similarly generated by the pursuit of efficiency. In their approach, alienability is prima facie correct or justified, and inalienability must be the exception that proves the rule. Their definition of inalienability collapses all inalienabilities into market-inalienability[55] by failing to distinguish between prohibiting all loss or transfer and prohibiting sale.

Using market methodology, Calabresi and Melamed argue that external costs might explain or justify inalienability. One category of external cost that might be prevented by inalienability is large-scale social cost that sellers can inflict on the public. Calabresi and Melamed use pollution as an example, but their reasoning could just as well apply to Saturday night specials, heroin, or cigarettes:

> For instance, if Taney were allowed to sell his land to Chase, a polluter, he would injure his neighbor Marshall by lowering the value of Marshall's land. Conceivably, Marshall could pay Taney not to sell his land; but, because there are many injured Marshalls, freeloader and information costs make such transactions practically impossible. . . . [W]here there are so many injured Marshalls that the price required under [a] liability rule is likely to be high enough so that no

---

[52] *See* Calabresi & Melamed, *supra* note 3, at 1111–15.

[53] *See, e.g.*, Epstein, *Why Restrain Alienation?*, 85 COLUM. L. REV. 970 (1985); Rose-Ackerman, *supra* note 10; *see also* pp. 1867–70.

[54] *See* Calabresi & Melamed, *supra* note 3, at 1106–10. Although they do not elaborate the point, Calabresi and Melamed think that the same regime is justified from a libertarian point of view. Property rules best satisfy libertarian concerns, because they generally require the least state intervention, but liability rules might serve libertarian interests better in certain circumstances, for example where property rules are especially difficult to enforce. *See id.* at 1092 n.7. Such a convergence of efficiency and liberty is often noted by those who tend toward universal commodification. *See supra* note 48.

[55] *See* Calabresi & Melamed, *supra* note 3, at 1092 ("An entitlement is inalienable to the extent that its transfer is not permitted between a willing buyer and a willing seller.").

> one would be willing to pay it, . . . [b]arring the sale to polluters will be the most efficient result because it is clear that avoiding pollution is cheaper than paying its costs — including its costs to the Marshalls.[56]

The argument displays a tendency toward universal commodification in two respects. First, its logic applies to gift transfers as well as sales, but only sales are mentioned, perhaps because all interactions between humans qualify as sales. Second, Calabresi and Melamed describe injury to third parties in market rhetoric; pollution harms people's land value, rather than their health and quality of life.

The other category of external cost that might be prevented by inalienability involves what Calabresi and Melamed call "moralisms." The term "moralism" refers to the assimilation of moral and political right to the market by conceiving of people's moral tenets as goods and assigning them a dollar value. This assimilation represents the ultimate reach of market rhetoric. For example:

> If Taney is allowed to sell himself into slavery, or to take undue risks of becoming penniless, or to sell a kidney, Marshall may be harmed, simply because Marshall is a sensitive man who is made unhappy by seeing slaves, paupers, or persons who die because they have sold a kidney. Again Marshall could pay Taney not to sell his freedom to Chase the slaveowner; but again, because Marshall is not one but many individuals, freeloader and information costs make such transactions practically impossible. . . . [And] since the external cost to Marshall does not lend itself to an acceptable objective measurement, . . . liability rules are [also] not appropriate.[57]

The authors refer to slavery, spendthrift trusts, and organ-selling, but could just as well have chosen child labor, gambling, or prostitution. This argument, too, evidences a tendency toward universal commodification. Because the argument logically prohibits gifts as well as sales, it may not capture the moral rejection of organ-selling. "Taney" could die just as well from giving away a kidney as from selling it.[58] On a deeper level, the argument disturbingly suggests that the inalienability rule against slavery would not be justified if the rule were inefficient. If enough of the "Marshalls" liked slavery, so that the

---

[56] *Id.* at 1111.

[57] *Id.* at 1112.

[58] If the authors mean that "Marshall" is made unhappy by "Taney's" death, and that this is a reason to make kidneys inalienable, they fail to recognize our moral approval of kidney-giving. If they mean that "Marshall" is made unhappy only by death after kidney-selling, on the other hand, and that this is a reason to make kidneys inalienable, they are postulating a "moralism" that distinguishes between gifts and sales in a way that market rhetoric cannot. Because they define inalienability as nonsalability, their theoretical apparatus cannot distinguish market-inalienability from other kinds, and hence ignores the moral distinction between gift and sale.

prohibition would be a cost rather than a benefit to them, slavery would be efficient and therefore (at least according to this argument) acceptable. This result is the triumph of market methodology.[59]

---

[59] In addition to the two categories of external costs, the reasons Calabresi and Melamed adduce for the existence of inalienabilty in an otherwise free market system include two categories of paternalism, and "distributional goals." *See id.* at 1113–15. Their paternalistic reasons for inalienability are characterized (surprisingly) as efficiency reasons. In "self-paternalism," the individual furthers her own long-run welfare maximization by forbidding herself certain contrary transactions in the short-run; the classic illustration of this is Ulysses tying himself to the mast to avoid succumbing to the Sirens. *See id.* at 1113. In "true paternalism," "the most efficient pie is no longer that which costless bargains would achieve, because a person may be better off if he is prohibited from bargaining." *Id.* at 1114. The examples the authors have in mind are prohibitions of certain activities by minors. They state that true paternalism, unlike the situations involving "moralisms," involves "the notion that at least in some situations the Marshalls know better than Taney what will make Taney better off." *Id.* at 1113. It is hard to understand how, consistent with the moral subjectivism revealed in the discussion of "moralisms" as external costs, the authors can make sense of the notion that the person under a prohibition would be "better off" in some view other than her own. Thus, their paternalism-efficiency argument may collapse into their moralism-externality argument. By imposing paternalistic restraints, we are benefiting those people whose subjective moral beliefs include the "knowledge" that others would be better off if restrained, and who attach subjective value to seeing them better off. (Perhaps the argument can be saved from this collapse by supposing that in appropriate situations, such as restraining minors, we can confidently predict that the person herself will eventually come to realize she is better off.) *See infra* note 189.

In showing how "distributional goals" bear on inalienability, Calabresi and Melamed suggest that we should be on guard against the "danger . . . that what is justified on, for example, paternalism grounds is really a hidden way of accruing distributional benefits for a group whom we would not otherwise wish to benefit." *Id.* at 1115. Thus, "prohibiting the sale of babies makes poorer those who can cheaply produce babies and richer those who through some nonmarket device get free an 'unwanted' baby." *Id.* at 1114. Although this argument is directed toward distribution rather than efficiency, it speaks in the rhetoric of universal commodification: one is "poorer" if she cannot sell a baby she can "cheaply produce."

The commitment to market rhetoric in fact seems to have made it difficult for Calabresi and Melamed to talk about "other justice reasons" relevant to entitlements:

> [W]e may as well admit that it is hard to know what content can be poured into ["other justice reasons"], at least given the very broad definitions of economic efficiency and distributional goals that we have used. . . . We defined distribution as covering *all* the reasons, other than efficiency, on the basis of which we might prefer to make Taney *wealthier* than Marshall. So defined, there obviously was no room for any other reasons.

Id. at 1102, 1104 (emphasis in original). Calabresi has since made it clear that he does not embrace universal market rhetoric:

> We could speak about all that we do in law using only "justice" or "rights" language. We could instead speak of all that we do using only "scientific policy-making" language, as if everything involved costs and benefits. Or we could use both languages . . . The question is, which language highlights the similarity among those things which we tend to trade off against each other readily . . . and separates out those things as to which we wish to make trade-offs only more rarely or perhaps not at all? . . . In other words, we should use the language which allows us to put together those things which we want to talk about together. . . . To translate from one language to another is to betray. That is the translation, and the betrayal, of an Italian saying: "*Traduttore, traditore.*" . . . What I am suggesting to you is that the use of economic language to describe part of law is terribly useful. And it is even useful, occasionally, to play as if one could use economic language across all of law, but only so long as one does not get confused about the fact that the real trade-offs in meaningful areas are not on a one to one basis. And that is

Later writers have essentially adopted the Calabresi-Melamed analysis. According to Richard Epstein, the only sound justification for inalienability is "the practical control of externalities."[60] As in Calabresi and Melamed's view, inalienability is the exception that proves the market rule;[61] it comes into being only to achieve what the market "would" achieve but cannot, because of various kinds of transaction costs.[62] Epstein's analysis of vote-selling as an externality problem reveals the scope of his market methodology and market rhetoric.[63] If an entrepreneur could buy the votes to put herself into public office, Epstein argues, she could then pay off the sellers with public money, thus depleting the common pool of assets for her own gain.[64] This argument relies on the universal commodification version

---

why, most of the time, it is better to limit the use of economic language to those issues in law where simple trade-offs *are* likely.

Calabresi, *Thoughts on the Future of Economics in Legal Education*, 33 J. LEGAL EDUC. 359, 363–64 (1983) (emphasis in original). It seems fair to infer that Calabresi would now disapprove of market rhetoric to consider the legal or moral treatment of baby-selling, as in the passage quoted above, or to consider the treatment of rape, which I discuss at pp. 1879–81 below.

[60] Epstein, *supra* note 53, at 990. According to Epstein, "[r]ules restraining alienation are best accounted for, both positively and normatively, by the need to control problems of external harm and the common pool." *Id.* at 970. Epstein's common pool argument is about costs that arise when a resource must be shared. It is in fact a variant of the tragedy of the commons. *See* Demsetz, *Toward a Theory of Property Rights*, 57 AM. ECON. REV. 347 (1967); Hardin, *The Tragedy of the Commons*, 162 SCIENCE 1243 (1968). Like the argument for the tragedy of the commons, Epstein's argument assumes that, absent restraint, people will maximize individual short-run gain to the ultimate degradation of a resource. *See* Epstein, *supra* note 53, at 978.

[61] *See* Epstein, *supra* note 53, at 971.

[62] Epstein's argument about external harm is akin to Calabresi and Melamed's arguments about external costs and efficiency-based paternalism. *See supra* note 59. The external harms he mentions fall into three categories: aggression against third parties, overexploitation of the common pool, and exploitation of infants and insane persons. The primary examples of inalienability Epstein has in mind are prohibitions: guns, liquor, and drugs. He speaks of them as restrictions or bans on sales, but the logic of the argument extends to gifts and to possession and use as well. The distinctions between market-inalienabilities and other kinds are not noticeable when everything is thought of as part of the market.

In Epstein's view, restraints voluntarily imposed by individual bargaining are presumptively efficient. Restraints imposed by law are to be regarded much more warily, but his common pool argument can justify a few of them. *See supra* note 60. As examples of common pool types of restraints imposed by law rather than individual bargaining, Epstein mentions water rights and voting rights in corporate and political elections. *See* Epstein, *supra* note 53, at 979–82, 984–88; *infra* note 64. Under universal commodification, of course, these are not qualitatively different kinds of "goods." Epstein sees the English common law of riparian rights, which tied water rights to land rights and limited water rights to uses that did not disturb the natural flow, as steering between two extremes: inefficiencies caused by free alienability (the tragedy of the commons), *see supra* note 60, and inefficiencies caused by permanent entitlement of users who do not value highly the resource (which might be called inefficiencies of status). *See* Epstein, *supra* note 53, at 981. In effect he is proposing that a properly tailored status-inalienability, *see supra* note 20, is a cure (or a palliative) for the tragedy of the commons.

[63] *See* Epstein, *supra* note 53, at 988.

[64] To see the extent of Epstein's market rhetoric, consider his opinion that the most likely motive for buying votes "is to obtain control of the public machinery, in ways that allow a

of interest-group pluralism, conceiving of politics as rent seeking by those who put their friends and sympathizers in office in order to line their own pockets.[65]

Although Epstein's theory purportedly rests on libertarian rights as well as economic efficiency,[66] it differs little from Calabresi and Melamed's. Epstein does not recognize distinctions between market-inalienability and other forms of inalienability, because for him the only real issue is whether a market is under the circumstances self-defeating so that market results must be achieved by other means. For him, the harms caused by treating rights of persons or citizens (such as voting) as alienable commodities are market types of harm — external costs.

Susan Rose-Ackerman, another scholar who carries forward the view of Calabresi and Melamed, finds three normative rationales for inalienabilities:[67] economic efficiency, "certain specialized distributive

---

person to recover, at the very least, the money that was paid out to the individuals who sold their votes, with something left to compensate the buyer for the labor and entrepreneurial risk." *Id.* at 987–88. Someone whose rhetoric is less thoroughly market-oriented might surely conceive the motive for buying votes to be advancing one's unmonetized political, social, religious, or moral ideas.

[65] *See supra* note 46.

[66] *See* Epstein, *supra* note 53, at 971. In earlier work, Epstein stressed libertarian rights. *See, e.g.*, Epstein, *Possession as the Root of Title*, 13 GA. L. REV. 1221 (1979). Recently he claimed that libertarian rights and utilitarian reasoning lead to the same institutional rules. *See, e.g.*, Epstein, *Past and Future: The Temporal Dimension in the Law of Property*, 64 WASH. U.L.Q. 667 (1986); Epstein, Remarks in *Time, Property Rights, and the Common Law* (Round Table Discussion), 64 WASH. U.L.Q. 793, 793 (1986) ("[M]y long-term campaign . . . is to explain why libertarian rules are the first approximation of a decent set of rights in the utilitarian world."). Robert Ellickson and I demonstrate that it is not as easy as Epstein claims to be simultaneously a libertarian and a utilitarian. *See* Ellickson, *Adverse Possession and Perpetuities Law: Two Dents in the Libertarian Model of Property Rights*, *supra* note 48, at 737; Radin, *Time, Possession and Alienation*, 64 WASH. U.L.Q. 739, 743–45 (1986). Most recently, Epstein seems to affirm that his foundational normative principle is indeed efficiency and not libertarian natural rights. He says, for example, that the traditionally recognized natural rights evolved instrumentally to serve efficiency before people were able to theorize explicitly about efficiency. *See* Remarks of Richard Epstein, in *Proceedings of the Conference on Takings of Property and the Constitution*, 41 U. MIAMI L. REV. 49, 125–27 (1986).

[67] *See* Rose-Ackerman, *supra* note 10, at 931. Rose-Ackerman recognizes that the Calabresi-Melamed definition of inalienability, *see supra* note 55, refers only to nonsalability (in my terms, market-inalienability), and she recognizes that the reason for this narrowness is that their "treatment of inalienability is colored by [the] emphasis on quid pro quo transfers," Rose-Ackerman, *supra* note 10, at 932, that forms the basis of their treatment of property rules and liability rules. Rose-Ackerman posits a many-celled matrix of inalienabilities, representing all possible combinations of three variables: limits on who may hold an entitlement; limits on uses of it (uses that are permitted, required, or forbidden); and limits on transfers (what kinds of transfers are permitted or forbidden). Even though she distinguishes 96 varieties of inalienability, her matrix does not distinguish among the various types of losses, such as abandonment, forfeiture, and cancellation, *see supra* pp. 1852–53, that inalienability might prevent. She appears to concentrate on transfers just as Calabresi and Melamed concentrate on quid pro quo transfers. Indeed, she gives the name "pure inalienability" to nontransferability. *See* Rose-

goals," and incompatibility of unfettered market processes with "the responsible functioning of a democratic state."[68] Unless one of these is implicated, unfettered market trading is presumptively desirable.[69] The efficiency rationale is a broadened transaction costs analysis, adding information and coordination problems to the more familiar externalities. Inalienabilities are "second-best responses" to these market failures.[70]

Rose-Ackerman does not present a normative framework for evaluating inalienabilities according to distributive justice. The rationale for inalienability based on distribution is "narrowly focused," referring to situations in which an inalienability can be used to single out recipients of a benefit.[71] More generally, Rose-Ackerman argues that using inalienability to achieve distributive goals is unjustified "except to prevent monopoly gains,"[72] and that "distributive costs" that arise when efficiency-based restrictions burden a particular group might render the restrictions unjustified.[73] Although her treatment of the distributive rationale thus seems undeveloped relative to efficiency,[74] Rose-Ackerman's is a hybrid analysis. By raising the issue of distributive justice and and by considering the incompatibility of market processes with democratic functioning, she means to combine economic analysis "with a sensitivity to noneconomic ideas."[75] Her argument connecting certain inalienabilities (such as voting) with ideals

---

Ackerman, *supra* note 10, at 935. The main problem with her approach, however, is a normative flatness. Because some types of inalienability that occupy cells in the matrix occur infrequently or not at all in our legal and moral culture, and others are very salient, the matrix tends to obscure the fact that some types of inalienability carry broad normative implications, whereas some are less significant and some are mere logical permutations. In her taxonomy, for example, "modified inalienability," in which sales are forbidden and gifts are permitted — market-inalienability — is the exact counterpart of "modified property," in which sales are permitted and gifts are forbidden. *See id.* at 935, 942–51. In addition, sixteen cells are occupied by various kinds of complete use restrictions, in which "nothing is permitted," and in four of these cases "no one" may hold the entitlement. *See id.* at 933–34.

[68] Rose-Ackerman, *supra* note 10, at 932–33.

[69] *See id.* at 932.

[70] *See id.* at 938.

[71] "If policymakers wish to benefit a particular sort of person but cannot easily identify these people ex ante, they may be able to impose restrictions on the entitlement that are less onerous for the worthy group than for others who are nominally eligible." *Id.* at 940. For example, Rose-Ackerman suggests that the Homesteading Acts can be justified as a means of transferring land to formerly landless people willing to live on the land and farm it. *See id.*

[72] *Id.* at 942, 948–49.

[73] *See id.* at 941.

[74] Rose-Ackerman's conclusion that "it is generally possible to conceive of an alternative policy that would be superior [to inalienability] if transaction costs were lower," *id.* at 969, seems to indicate that efficiency — even if it is efficiency in achieving aims that are "distributive" — is her main concern. To this conclusion there is a "major exception" involving the "ideal of citizenship, where insulation from market forces may be desirable in principle." *Id.*

[75] *Id.* at 931.

of citizenship espouses a kind of pluralism, a separation of politics and markets.[76] Nevertheless, it is couched almost exclusively in market rhetoric.[77]

In order to evaluate pluralist positions that are not pluralist in rhetoric, it is necessary to consider the normative role of market rhetoric: is commodification in rhetoric tantamount to commodification?[78] As background for considering that question, I turn now to the critique of universal commodification.

III. THE CRITIQUE OF UNIVERSAL COMMODIFICATION

A traditional critical response to universal commodification, at least since Marx, has been a global rejection of commodification. Universal decommodification or noncommodification maintains that the market ought not to exist and that social interactions involving production and consumption should be reconceived in a nonmarket way. Even if one rejects that ideal, however, as I do because of a problem of transition, the critique of universal commodification offers a crucial insight: a world in which human interactions are conceived of as market trades is different from one in which they are not. Rhetoric is not just shaped by, but shapes, reality.

---

[76] *See supra* note 74.

[77] Rose-Ackerman describes four conceptions of the duties of citizens, each associated with a different form of inalienability. *See* Rose-Ackerman, *supra* note 10, at 961–68. Under the "strong" conception, represented by inalienable duties, "the state requires certain actions of some or all citizens and forbids the transfer of these duties to others." *Id.* at 966. Under the "weakest" conception, represented by alienable property rights, "citizenship services" are obtained under a market regime. Votes can be sold or given away, and jurors and soldiers are paid volunteers. *See id.* at 962. Under inalienable rights, citizenship rights (like voting) are nontransferable, but there is no duty to exercise them. *See id.* at 966. Under alienable duties, "people are assigned duties by the state" but can pay or persuade others to perform them. *Id.* at 967. Rose-Ackerman hints that she would consider distribution as well as efficiency in selecting one of these conceptions of citizenship, noting that it would be unacceptable to have the rich end up with most of the votes. *See id.* at 963. She hints at an ethical case for the inalienable duty conception of citizenship, but she puts it in market rhetoric: this conception of the military draft, or jury service, treats these duties as responsibilities of citizenship, but "it is clearly more costly for society as a whole since some people with high opportunity wages in the private sector . . . will be called." *Id.* at 967. In general, Rose-Ackerman's matrix demonstrates the liberal positivist conception of the state as creating or granting whatever rights and duties citizens have. Under all "conceptions" of citizenship the state and its citizens are radically separate. In my view, her adherence to market rhetoric, her failure to provide a normative structure for determining when distribution trumps efficiency, and her individualist positivism combine to weaken her third normative rationale for inalienability, in which she finds market processes sometimes incompatible with democratic functioning. Nonetheless, she raises a question that is important to pursue. I believe further investigation of it must explore the notion of community-inalienability. *See supra* note 21.

[78] This question is taken up in Section B of Part III below.

A. Universal Noncommodification

Universal noncommodification holds that the hegemony of profit-maximizing buying and selling stifles the individual and social potential of human beings through its organization of production, distribution, and consumption, and through its concomitant creation and maintenance of the person as a self-aggrandizing profit- and preference-maximizer. Anticommodifiers tend to assume that we are living under a regime of universal commodification, with its attendant full-blown market methodology and market rhetoric. They also tend to assume that universal commodification is a necessary concomitant of commodification in the narrower sense — the existence of market transactions under capitalism. Anticommodifiers link rhetoric and reality in their assumption that our material relationships of production and exchange are interwoven with our discourse and our understanding of ourselves and the world.[79]

1. Alienability and Alienation: The Problem of Fetishism. — For critics of the market society, commodification simultaneously expresses and creates alienation. The word "alienation" thus harbors an ironic double meaning. Freedom of alienation is the paramount characteristic of liberal property rights, yet Marx saw a necessary connection between this market alienability and human alienation. In his early writings, Marx analyzed the connection between alienation and commodity production in terms of estranged labor; later he introduced the notion of commodity fetishism.[80] In his treatment of estranged labor, Marx portrayed workers' alienation from their own human self-activity as the result of producing objects that became market commodities. By objectifying the labor of the worker, commodities create object-bondage and alienate workers from the natural world in and with which they should constitute themselves by creative interaction. Ultimately, laboring to produce commodities turns the worker from a human being into a commodity, "indeed the most wretched of commodities."[81] Marx continued:

> The worker becomes an ever cheaper commodity the more commodities he creates. With the *increasing value* of the world of things

[79] This is the link between rhetoric and the world that I suggest (in Section B of this Part) can be taken seriously even outside the context of the worldview of complete noncommodification. The suggestion that market transactions (for example, in sexual services or children) might engender a domino effect in rhetoric that leaves everyone unable to conceive of possible non-market alternatives is one way this normative view of rhetoric will figure in my evolutionary pluralist view. *See infra* pp. 1912–14.

[80] For discussion of Marx's early and later treatment of alienation and its relationship to commodification, see S. AVINERI, THE SOCIAL AND POLITICAL THOUGHT OF KARL MARX 96–123 (1968), and A. GOULDNER, THE TWO MARXISMS 177–220 (1980).

[81] Marx, *Economic and Philosophic Manuscripts of 1844*, in THE MARX-ENGELS READER 70 (R. Tucker 2d ed. 1984) [hereinafter THE MARX-ENGELS READER].

> proceeds in direct proportion the *devaluation* of the world of men. Labour produces not only commodities; it produces itself and the worker as a *commodity* — and does so in the proportion in which it produces commodities generally.[82]

Commodification brings about an inferior form of human life.[83] As a result of this debasement, Marx concluded that people themselves, not just their institutions, must change in order to live without the market. To reach the post-capitalist stage, "the alteration of men on a mass scale is necessary."[84]

The fetishism of commodities represents a different kind of human subjection to commodities (or a different way of looking at human subjection to commodities).[85] By fetishism Marx meant a kind of projection of power and action onto commodities. This projection

---

[82] *Id.* at 71 (emphasis in original).

[83] This Marxist view should be contrasted with the traditional view that the market brings about better human beings. Albert O. Hirschman reviews the strand of traditional thought holding that the growth of the market was a civilizing and humanizing force in society — the "*doux-commerce*" thesis prevalent in the eighteenth century — and contrasts it with other traditional views. *See* Hirschman, *Rival Interpretations of Market Society: Civilizing, Destructive, or Feeble?*, 20 J. ECON. LITERATURE 1463, 1483 (1982). In the current trend toward recognizing the importance of long-term relational contracts, Hirschman detects the possibility that "[t]he stage could thus be best set for a partial rehabilitation of the '*doux-commerce*' thesis," *id.* at 1474, whereas anticommodifiers or critical pluralists would see instead an evolutionary trend toward decommodification, *cf.* Macneil, *Relational Contract: What We Do and Do Not Know*, 1985 WIS. L. REV. 483; Gordon, *Macaulay, Macneil, and the Discovery of Solidarity and Power in Contract Law*, 1985 WIS. L. REV. 565.

[84] Marx, *The German Ideology: Part I*, in THE MARX-ENGELS READER, *supra* note 81, at 193. The revolution is necessary "not only because the *ruling* class cannot be overthrown in any other way, but also because the class *overthrowing* it can only in a revolution succeed in ridding itself of all the muck of ages and become fitted to found society anew." *Id.* (emphasis in original).

[85] *See* 1 K. MARX, CAPITAL, 71–83 (F. Engels ed. 1894, S. Moore & E. Aveling trans. 1984); *cf.* Balbus, *Commodity Form and Legal Form: An Essay on the "Relative Autonomy of the Law"*, 11 LAW & SOC'Y REV. 571, 573–75 (1977) (expounding Marx's theory of the fetishism of commodities).

Thoroughgoing reification, with its ramifications for the disempowerment of human beings, is the classical meaning of commodity fetishism. *See infra* pp. 1873–74. There are other meanings as well. One refers to the surface phenomenon of rampant consumerism or crass devotion to material possessions. To be a commodity fetishist in this newer and less technical sense is simply to have one's identity too tied to possessions, to be too dependent upon thing-ownership for pleasure and a sense of self-worth. This meaning does not correspond to commodity fetishism in the classical Marxist sense because it does not refer specifically to the nature of the things possessed as capitalist market trade artifacts. Nevertheless, it is a form of fetishism (projection onto objects), and it is certainly compatible with some Marxist views of the world. Another meaning is what Fred Hirsch calls "the new commodity fetishism," the idea that "an excessive proportion of individual activity is channeled through the market so that the commercialized sector of our lives is unduly large." F. HIRSCH, SOCIAL LIMITS TO GROWTH 84 (1976). This "new commodity fetishism" sticks to the technical meaning of "commodity" and is thus different from the common-sense view I have just described, which is more truly "fetishist." It is also apparently pluralist, not Marxist.

reflects — but disguises — human social interactions. Relationships between people are disguised as relationships between commodities, which appear to be governed by abstract market forces. I do not decide what objects to produce, rather "the market" does. Unless there is a demand for paperweights, they will have no market value, and I cannot produce them for sale. Moreover, I do not decide what price to sell them for, "the market" does. At market equilibrium, I cannot charge more nor less than my opportunity costs of production without going out of business. In disequilibrium, my price and profit are still set by "the market"; my price depends upon how many of us are supplying paperweights in relation to how many people want to buy them and what they are willing to pay for them. Thus, the market value[86] of my commodity dictates my actions, or so it seems. As Marx put it, "[producers'] own social action takes the form of the action of objects, which rule the producers instead of being ruled by them."[87]

In an analysis that has profoundly influenced many contemporary anticommodifiers, Georg Lukács, developing Marx's concept of commodity fetishism, found commodification to be "the central, structural problem of capitalist society in all its aspects."[88] Lukács linked the trend to commodify the worker with Weberian "rationalization" of the capitalist structure.[89] The more efficient production becomes, the more fungible are the laborers. Moreover, fungibility becomes pervasive:

> [T]he principle of rational mechanisation and calculability [embraces] every aspect of life. Consumer articles no longer appear as the products of an organic process within a community (as for example in a village community). They now appear, on the one hand, as abstract members of a species identical by definition with its other members

---

[86] What we now call market value, Marx thought of as "exchange value," which he contrasted with "use value" (the worth of something to consumers) and "value" (the amount of labor socially necessary to produce something). *See* 1 K. MARX, *supra* note 85, at 84–93. For explication and criticism of Marx's theories of value, see, for example, Cohen, *Labor, Leisure, and a Distinctive Contradiction of Advanced Capitalism*, in MARKETS AND MORALS, cited in note 29 above, at 107, and J. ELSTER, MAKING SENSE OF MARX 119–65 (1985).

[87] 1 K. MARX, *supra* note 85, at 79. *cf.* Baker, *Property and Its Relation to Constitutionally Protected Liberty*, 134 U. PA. L. REV. 741 (1986) (arguing that "market oriented liberty" — as opposed to individual liberty defined as self-determination and self-realization — is not conducive to the autonomy of either producers or consumers).

[88] G. LUKÁCS, *Reification and the Consciousness of the Proletariat*, in HISTORY AND CLASS-CONSCIOUSNESS 83 (R. Livingstone trans. 1971). Lukács thought it might justifiably be claimed that Marx's "chapter dealing with the fetish character of the commodity contains within itself the whole of historical materialism," *id.* at 170, and that we can "gain an understanding of the whole of bourgeois society from its commodity structure," *id.* at 198.

[89] *See id.* at 88–92. "Rationalization" is Max Weber's term for the development of the economic system toward achieving ever greater profit at less cost. *See, e.g.*, A. KRONMAN, MAX WEBER 130–37 (discussing the formal rationality of economic action).

and, on the other hand, as isolated objects the possession or nonpossession of which depends on rational calculations.[90]

These falsely objectified commodities are said to be reified. According to Lukács, reification penetrates every level of intellectual and social life.[91] False objectification — false separateness from us — in the way we conceive of our social activities and environment reflects and creates dehumanization and powerlessness. The rhetoric, the discourse in which we conceive of our world, affects what we are and what our world is. For example, Lukács thought that the universal commodification of fully developed capitalism underlies physicalist reductionism in science and the tendency to conceive of matter as external and real.[92] He thought that universal commodification also underlies both our rigid division of the world into subjects versus objects ("the metaphysical dilemma of the relation between 'mind' and 'matter'"),[93] and the "Kantian dilemma" that places objective reason, purportedly the foundation of metaphysics and ethics, in the noumenal realm forever beyond our reach.[94] For Lukács, thought and reality are inextricably linked.[95]

---

[90] G. LUKÁCS, *supra* note 88, at 91.

[91] Lukács argued:

> The transformation of the commodity relation into a thing of "ghostly objectivity" cannot therefore content itself with the reduction of all objects for the gratification of human needs to commodities. It stamps its imprint upon the whole consciousness of man; his qualities and abilities are no longer an organic part of his personality, they are things which he can "own" or "dispose of" like the various objects of the external world. And there is no natural form in which human relations can be cast, no way in which man can bring his physical and psychic "qualities" into play without their being subjected increasingly to this reifying process.

Id. at 100.

[92] In reified bourgeois thought, "facts" are the highest form of fetishism:

> [I]n the "facts" we find the crytallisation of the essence of capitalist development into an ossified, impenetrable thing alienated from man. And the form assumed by this ossification and this alienation converts it into a foundation of reality and of philosophy that is perfectly self-evident and immune from every doubt. . . . Thus only when the theoretical primacy of the "facts" has been broken, only when *every phenomenon is recognised to be a process*, will it be understood that . . . the facts are nothing but the parts, the *aspects* of the total process that have been broken off, artificially isolated and ossified.

Id. at 184 (emphasis in original).

[93] *Id.* at 116.

[94] *See id.* at 114–17.

[95] "Only by conceiving of thought as a form of reality, as a factor in the total process can philosophy overcome its own rigidity dialectically and take on the quality of Becoming." *Id.* at 203 (footnote omitted). Lukács warned against conceiving of the link between thought and reality in a way that reintroduces foundationalism:

> It is true that reality is the criterion for the correctness of thought. But reality is not, it becomes — and to become the participation of thought is needed. . . .
>
> Thus thought and existence are not identical in the sense that they "correspond" to each other, or "reflect" each other, that they "run parallel" to each other or "coincide" with each other (all expressions that conceal a rigid duality). Their identity is that they are aspects of one and the same real historical and dialectical process.

Id. at 204.

2. Inalienability and Noncommodification: The Problem of Transition. — Earlier I noted that market-inalienability does not exist as a separate category for universal commodifiers, because nonsalability by definition encompasses the universe of inalienabilities.[96] Likewise, universal noncommodifiers do not distinguish market-inalienability as an analytical category. Market-inalienability posits a nonmarket realm that appropriately coexists with a market realm, and this implicitly grants some legitimacy to market transactions, contrary to the noncommodifier's premise. Thus, only those who think that market and nonmarket realms legitimately coexist — pluralists — readily recognize market-inalienability.

Nevertheless, some who espouse universal noncommodification for the long run might espouse pluralism in the short run, if they think that introducing piecemeal market-inalienabilities is a way of making progress toward universal noncommodification. True utopian noncommodifiers, however, would oppose even this interim pluralism; for them, inalienability should be eschewed because it recognizes the legitimacy of alienability, the heart of capitalist property relationships.

I shall call these two approaches to noncommodification evolutionary and revolutionary. The revolutionary approach criticizes, as misguided and as an artifact of capitalism, the entire world view that posits a structure of persons versus objects, and alienable versus inalienable objects. It holds that the capitalist structure permeates not only our world of social interaction and allocation of resources but also our discourse, vocabulary, and conception of human flourishing. By contrast, the evolutionary approach, interim pluralism, recognizes the necessity of working within existing market structures of capitalism to achieve universal noncommodification. It differs from the other pluralist views that seek to curtail the scope of the market only in that it does not condone the remaining market order after piecemeal inalienabilities are in place. These two approaches exemplify a pervasive dilemma for social progress: whether and how existing conceptions and structures, such as commodification, can be used now to ensure they will no longer be used in some better future. This is the problem of transition.[97]

The evolutionary approach harbors a transition problem because it does not address how we can progress toward noncommodification using existing social structures and conceptual schemes that are thought to be artifacts of commodification. Partial decommodification

---

[96] *See supra* p. 1863.

[97] If the means-ends distinction is denied, the transition never ends. Transition, or the seeking of ideals that we formulate and yet see as beyond us, is then the central feature of political life. *See, e.g.*, R. UNGER, KNOWLEDGE AND POLITICS (1975); Baker, *The Process of Change and the Liberty Theory of the First Amendment*, 55 S. CAL. L. REV. 293 (1982); Cover, *The Supreme Court, 1982 Term — Foreword:* Nomos *and Narrative*, 97 HARV. L. REV. 4 (1983).

in the context of a continuing implicit commitment to a dominant market order may mean that any deviations from the market order will only reinforce commodification, by being seen merely as exceptions that prove the market rule. The revolutionary approach also harbors a transition problem because it does not — indeed, cannot — suppose that we shall somehow arrive all at once in the promised land of total noncommodification. Yet if radical decommodification is attempted for less than everything, it appears evolutionary, not revolutionary, and it may wreak injustice.

The problem of transition for revolutionary decommodification can be illustrated by examining a universal decommodification argument regarding our system of damage remedies in tort law.[98] Richard Abel advocates replacing the tort system with a system that treats people equally, regardless of whether their misfortunes are caused by their own fault, other people's fault, an unavoidable accident, an illness, or a congenital disability.[99] The system should not compensate for damages to property or individual earning power, because such compensation reaffirms and maintains inequality.[100] In addition, it should not compensate for intangible injuries, because this contributes to a cultural view of experience and love as commodities. Damages for pain and suffering "commodify our unique experience;" damages for injuries to relationships, such as loss of consortium or witnessing the injury to a loved one, "commodif[y] love."[101] Abel explains:

> Just as society pays pain and suffering damages to the injured victim who is shunned (so he can purchase the commodified care and companionship that will no longer be given out of love and obligation), so it pays damages to those who loved him, compensating them for their lost "investment" in the relationship (so they can invest in other human capital).[102]

According to Abel, we should not assume that people are willing to undergo suffering or loss for a sum of money, because we should

---

[98] The problem of transition from an evolutionary point of view will be considered in Part IV when I turn to an examination of pluralism.

[99] *See* Abel, *A Critique of American Tort Law*, 8 BRITISH J. L. & SOC'Y 199, 207–09 (1981); Abel, *A Socialist Approach to Risk*, 41 MD. L. REV. 695 (1982); Abel, *Torts*, in THE POLITICS OF LAW 185 (D. Kairys ed. 1982).

[100] *See* Abel, *A Critique of American Tort Law*, *supra* note 99, at 201–02.

[101] *Id.* at 207. For a case that commodifies more explicitly than most, see Sherrod v. Berry, 629 F. Supp. 159, 163 (N.D. Ill. 1985), in which an economist was allowed to testify that the hedonic value of a human life is "somewhere in the dimension of three times up to 30 times [the person's] economic productive income." *Cf.* L. HYDE, THE GIFT: IMAGINATION AND THE EROTIC LIFE OF PROPERTY 62–66 (discussing the Pinto case, in which Ford Motor Company used cost-benefit analysis to weigh the costs of injury and death against the costs of an $11 safety device that would have prevented the Pintos' gas tanks from exploding).

[102] Abel, *A Critique of American Tort Law*, *supra* note 99, at 208 (footnote omitted).

not assume that these capacities are alienable.[103] The assumption is dehumanizing and (if one is true to Marx) a self-fulfilling prophecy, bringing about a human world in which people really are estranged from their essential human capacities. By refusing to allow recovery for these kinds of injuries, we would be saying that human life activity, or at least certain aspects of it, ought not to be traded, nor to be conceived of in market rhetoric or evaluated in market methodology. Abel's system is thus a revolutionary proposal to decommodify the law of personal injury.

Many people, even some who deplore commodification, will find the proposal troubling and its agenda unjust. To deny money damages, inadequate though they may be, seems to compound the injury to tort victims under the present social structure, in which we have not put into practice other measures that would take care of them in better ways or prevent their injuries in the first place. This piecemeal decommodification appears unjust in its unredeemed context because of the pervasive dilemma of transition. Abel argues that existing conceptions and structures of tort law cannot be used if we are to achieve a more humane social order, but it also seems unjust not to use them during the transition to the imagined better world. This central dilemma of social change recurs in many contexts involving decommodification. Failure to face it satisfactorily is the primary shortcoming of both revolutionary and evolutionary arguments for universal noncommodification. Revolutionary noncommodification might wreak great injustice and interim pluralism might make no progress.

B. The Moral and Political Role of Rhetoric

Although the problem of transition gives us reason not to accept arguments for universal noncommodification, an implicit but central philosophical commitment of many universal noncommodifiers should be embraced because it plays a necessary role in pluralism as I believe it must now be reconceived. That commitment is the view that our discourse and our reality are interdependent.

"The word is not the thing," we were taught, when I was growing up.[104] Rhetoric is not reality; discourse is not the world. Why should it matter if someone conceptualizes the entire human universe as one giant bundle of scarce goods subject to free alienation by contract, especially if reasoning in market rhetoric can reach the same result

---

[103] *See id.* at 207.

[104] This slogan seems to have been popularized by Samuel Hayakawa. *See* S. HAYAKAWA, LANGUAGE IN THOUGHT AND ACTION 24–25 (4th ed. 1978). It, along with a companion "extensionalist" slogan, "A map is not the territory," apparently stem from the work of Alfred Korzybski. *See, e.g.*, A. KORZYBSKI, SCIENCE AND SANITY 750 (4th ed. 1958).

that some other kind of normative reasoning reaches on other grounds? Consider three possible answers: it matters because the rhetoric might lead less-than-perfect practitioners to wrong answers in sensitive cases; it matters because the rhetoric itself is insulting or injures personhood regardless of the result; or — the implicit philosophical commitment of the anticommodifiers — it matters because *there is no such thing as two radically different normative discourses reaching the "same" result.*

1. Risk of Error. — The rhetoric of commodification might lead imperfect practitioners to wrong answers, even if the sophisticated practitioner would not be misled. In other words, commodification-talk creates a serious risk of error in certain cases. To see this, it may be helpful to compare a normative heuristic like cost-benefit analysis to a flat map of the world. Such a map is easy to use at the point of projection, but difficult and misleading at the edges.[105] Cost-benefit analysis is not difficult when two firms deal with each other, if we define firms as profit-maximizing black boxes and no difficult externalities exist. By contrast, cost-benefit analysis involving people's subjective well-being is difficult to get right when many different people are involved and we are talking about interests they hold dear.[106] For example, the economic analysis of residential rent control could take into account not only the monetary costs to landlords and would-be tenants, but also the decline in well-being of tenants who are forced to lose their homes, break up their communities, and endure the frustration, disruption, and other "costs" of moving. But in practice the analysis proceeds differently.[107] Reasoning in market rhetoric, with its characterization of everything that people value as monetizable and fungible, tends to make it easy to ignore these other "costs." Money costs and easily monetizable matters are at the center of the map, and personal and community disruption are at the edges. Because it tends to ignore "costs" that are not readily monetizable, commodification-talk tends to err on the side of alienation.

---

[105] A similar point is made by Brian Barry in his discussion of what he perceives to be a shift from consequentialist to nonconsequentialist theories of ethics. "In principle, it is possible to imagine that you could reach the same concrete judgments starting from either point, but in practice it tends to make a big difference which cases you take as clear and central and which as difficult and exceptional." Barry, Book Review, 88 YALE L.J. 629, 630 (1979). Although Barry does not give illustrations, an example might be the justification of punishment. For utilitarian deterrence theorists, deciding whether to punish strict liability or *malum prohibitum* offenses does not cause problems, but deciding whether to punish innocent people or undeterrable offenses requires fancy footwork. For deontological retributivist theorists, the situation is reversed.

[106] *See, e.g.*, Baker, *supra* note 44; Kennedy, *supra* note 44.

[107] For example, a typical economic analysis can be found in Hirsch, *From "Food for Thought" to "Empirical Evidence" About Consequences of Landlord-Tenant Laws*, 69 CORNELL L. REV. 604 (1984).

2. Injury to Personhood. — In some cases market discourse itself might be antagonistic to interests of personhood. Recall that Posner conceives of rape in terms of a marriage and sex market.[108] Posner concludes that "the prevention of rape is essential to protect the marriage market . . . and more generally to secure property rights in women's persons."[109] Calabresi and Melamed also use market rhetoric to discuss rape.[110] In keeping with their view that "property rules" are prima facie more efficient than "liability rules" for all entitlements, they argue that people should hold a "property rule" entitlement in their own bodily integrity.[111] Further, they explain criminal punishment by the need for an "indefinable kicker," an extra cost to the rapist "which represents society's need to keep all property rules from being changed at will into liability rules."[112] Unlike Posner's view, Calabresi and Melamed's can be understood as pluralist,[113] but like

---

[108] *See supra* note 50 and accompanying text.

[109] R. POSNER, *supra* note 8, at 202. In the passage in which this sentence appears, Posner examines the argument that rape should not be punished criminally if there is "no market substitute for rape" because the rapist "derives extra pleasure from the coercive character of his act." (Presumably, the "market substitutes" would be marriage, dating, and prostitution.) Posner finds the argument "weak" — and is thus able to conclude that rape should be punished criminally — for three reasons: protecting the marriage market and property rights in women's persons; avoiding "wasteful expenditures" on protecting women and on overcoming the protections; and "the fact that the rapist cannot find a consensual substitute does not mean that he values the rape more than the victim disvalues it." *Id.*

[110] *See* Calabresi & Melamed, *supra* note 3, at 1124–27 (applying their framework to criminal sanctions).

[111] *See id.* at 1125–27.

[112] *Id.* at 1126.

[113] Although Calabresi and Melamed have a strong tendency to talk in monetized efficiency terms, there is a hint of pluralism in this passage, which must be quoted at some length in order to convey the rhetorical flavor:

> The question remains, however, why *not* convert all property rules into liability rules? The answer is, of course, obvious. Liability rules represent only an approximation of the value of the object to its original owner and willingness to pay such an approximate value is no indication that it is worth more to the thief than to the owner. . . . If this is so with property, it is all the more so with bodily integrity, and we would not presume collectively and objectively to value the cost of a rape to a victim against the benefit to the rapist even if economic efficiency is our sole motive. Indeed when we approach bodily integrity we are getting close to areas where we do not let the entitlement be sold at all and where economic efficiency enters in, if at all, in a more complex way. . . .
>
> The first year student might push on, however, and ask why we treat the thief or rapist differently from the injurer in an auto accident or the polluter in a nuisance case. Why do we allow liability rules there? In a sense, we have already answered the question. The only level at which, before the accident, the driver can negotiate for the value of what he might take from his potential victim is one at which transactions are too costly. The thief or rapist, on the other hand, could have negotiated without undue expense (at least if the good was one which we allowed to be sold at all) because we assume he knew what he was going to do and to whom he would do it.

Id. at 1125–27 (emphasis in original). Recall that Calabresi and Melamed also hint at pluralism in their mention of "other justice reasons" for setting entitlements, but they find it difficult to flesh out this idea, in my view because of their commitment to market rhetoric in that article. It should be noted that Calabresi has modified his views and probably no longer conceives of rape in market rhetoric. *See supra* note 59.

Posner's, their view conceives of rape in market rhetoric. Bodily integrity is an owned object with a price.[114]

What is wrong with this rhetoric? The risk-of-error argument discussed above is one answer. Unsophisticated practitioners of cost-benefit analysis might tend to undervalue the "costs" of rape to the victims. But this answer does not exhaust the problem. Rather, for all but the deepest enthusiast, market rhetoric seems intuitively out of place here, so inappropriate that it is either silly or somehow insulting to the value being discussed.

One basis for this intuition is that market rhetoric conceives of bodily integrity as a fungible object.[115] A fungible object is replaceable with money or other objects; in fact, possessing a fungible object is the same as possessing money. A fungible object can pass in and out of the person's possession without effect on the person as long as its market equivalent is given in exchange.[116] To speak of personal attributes as fungible objects — alienable "goods" — is intuitively wrong. Thinking of rape in market rhetoric implicitly conceives of as fungible something that we know to be personal, in fact conceives of as fungible property something we know to be too personal even to be personal property.[117] Bodily integrity is an attribute and not an

---

[114] Calabresi and Melamed's discussion of crimes against property and bodily integrity conceives of them as exactly parallel but for the concession quoted in note 113 above. Note also that a tendency toward universal commodification results from their use of market rhetoric: all scarce and desired resources are "goods." Calabresi and Melamed argue that "an entitlement to a good or to its converse is essentially inevitable. . . . We either have the right to our own property or body or the right to share others' property or body. We may buy or sell ourselves into the opposite position, but we must start somewhere." Calabresi & Melamed, *supra* note 3, at 1100–01 (footnotes omitted). And all these goods seem to have associated monetized costs and benefits: "Any entitlement given away free implies a converse which must be paid for. For all those who like children, there are those who are disturbed by children; for all those who detest armies, there are those who want what armies accomplish." *Id.* at 1099 n.23.

[115] In Radin, *Property and Personhood*, 34 STAN. L. REV. 957 (1982), I suggest that property may be divided into fungible and personal categories for purposes of moral evaluation. Property is personal in a philosophical sense when it has become identified with a person, with her self-constitution and self-development in the context of her environment. Personal property cannot be taken away and replaced with money or other things without harm to the person — to her identity and existence. In a sense, personal property becomes a personal attribute. On the other hand, property is fungible when there is no such personal attachment. *See id.* at 959–61, 978–79, 986–88.

[116] Thus, fungible objects are commodified: trading them is like trading money. Personal things are not commodified (or have been decommodified by assimilation into the person); the effect of detaching them from the person is nonmonetizable.

[117] The distinction between fungible and personal property is intended to distinguish between, on the one hand, things that are really "objects" in the sense of being "outside" the person, indifferent to personal constitution and continuity, and on the other hand, things that have become at least partly "inside" the person, involved with one's continuing personhood. The traditional subject/object dichotomy makes the notion of personal property hard to grasp, *see infra* pp. 1892–98, and, in the present context, poses a danger. To analogize bodily integrity to

object. We feel discomfort or even insult, and we fear degradation or even loss of the value involved, when bodily integrity is conceived of as a fungible object.

Systematically conceiving of personal attributes as fungible objects is threatening to personhood, because it detaches from the person that which is integral to the person. Such a conception makes actual loss of the attribute easier to countenance. For someone who conceives bodily integrity as "detached," the same person will remain even if bodily integrity is lost; but if bodily integrity cannot be detached, the person cannot remain the same after loss.[118] Moreover, if my bodily integrity is an integral personal attribute, not a detachable object, then hypothetically valuing my bodily integrity in money is not far removed from valuing *me* in money. For all but the universal commodifier, that is inappropriate treatment of a person.[119]

3. The "Texture of the Human World." — The difference between conceiving of bodily integrity as a detached, monetizable object and finding that it is "in fact" detached is not great, because there is no bright line separating words and facts. The modern philosophical turn toward coherence or antifoundationalist theories[120] means that

---

personal property may simply reintroduce the suggestion inherent in market rhetoric that I am trying to argue against: the suggestion that bodily integrity is somehow an owned object separate from personhood, rather than an inseparable attribute of personhood.

[118] This should not be understood to argue that someone who is raped is changed into a completely different person. To assert either that she is altogether the "same" afterwards or that she is completely "different" afterwards would trivialize her experience: we must have a way of conceptualizing our understanding both that she is different afterwards, so that we recognize that she has been changed by the experience, and simultaneously that she is the same afterwards, or else there would be no "she" that we can recognize to have had the experience and been changed by it. Just as personal attributes should not be seen as separate from an abstract self, neither should our experiences be seen as separate from ourselves.

[119] The intuitive view outlined here can be deepened by an understanding of the unsatisfactory view of personhood underlying universal commodification, and by an understanding of the role of the subject/object dichotomy in the ideological heritage of liberal pluralism. After considering these matters, *see infra* pp. 1884–85, 1892–98, I return to the problem of what is wrong with conceiving of bodily integrity in market rhetoric at pp. 1907–09 below.

[120] Antifoundationalism denies that rationality or truth consists of linear deductions from an unquestioned foundational reality or truth. Coherence theories stress holistic interdependence of an entire body of beliefs and commitments; they judge truth or rightness by fit, not by correspondence with an external foundational standard. For example, Rawls's "reflective equilibrium" is a moral methodology based on coherence, *see* J. RAWLS, A THEORY OF JUSTICE 48–51 (1971); and Quine's "field of force" is a metaphor for the coherence view of metaphysics, *see* W.V.O. QUINE, *Two Dogmas of Empiricism*, in FROM A LOGICAL POINT OF VIEW 20, 42 (2d ed. 1980). As Quine recognized, the coherence view tends toward pragmatism. *See id.* at 42–46. In my view, the need to reevaluate reality and truth in light of the rejection of foundationalism stems from the revolution wrought by Wittgenstein and Kuhn. *See* L. WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS (3d ed. 1958); T. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (2d ed. 1970). In a famous example, Kuhn suggests that "the scientist who looks at a swinging stone can have no experience that is in principle more elementary than seeing a

we cannot be sanguine about radically different normative discourses reaching the "same" result. Even if everybody agrees that rape should be punished criminally, the normative discourse that conceives of bodily integrity as detached and monetizable does not reach the "same" result as the normative discourse that conceives of bodily integrity as an integral personal attribute. If we accept the gist of the coherence or antifoundationalist theories, facts are not "out there" waiting to be described by a discourse. Facts are theory-dependent and value-dependent. Theories are formed in words. Fact- and value-commitments are present in the language we use to reason and describe, and they shape our reasoning and description, and the shape (for us) of reality itself.[121] Hilary Putnam's striking parable of the super-Benthamites illustrates how a view of values can alter one's view of the facts, the discourse in which one conceives and describes both fact and value, and thus the human world.[122] Putnam asks us to suppose that the continent of Australia is inhabited by people whose sole ethical imperative is that one should always act to maximize "hedonic tone."[123] Because they are single-minded, these people would do what appears to us to be ruthless:

---

pendulum. The alternative is not some hypothetical 'fixed' vision, but vision through another paradigm, one which makes the swinging stone something else." *Id.* at 128.

Hermeneutics, the sociology of knowledge, and perhaps post-structuralism are other thought traditions that converge with coherence theory on the issue of the theory- or discourse-dependence of reality. *See* P. BERGER & T. LUCKMANN, THE SOCIAL CONSTRUCTION OF REALITY (1966); J. CULLER, ON DECONSTRUCTION: THEORY AND CRITICISM AFTER STRUCTURALISM (1982); H. GADAMER, TRUTH AND METHOD (2d ed. 1982); HERMENEUTICS AND PRAXIS (R. Hollinger ed. 1985); C. NORRIS, DECONSTRUCTION: THEORY AND PRACTICE (1982); R. RORTY, PHILOSOPHY AND THE MIRROR OF NATURE (1979); Heller, *Structuralism and Critique*, 36 STAN. L. REV. 127 (1984); Peller, *The Metaphysics of American Law*, 73 CALIF. L. REV. 1152 (1985). These thought traditions diverge on the issue of whether discourses can be judged as better or worse, which is one reason I pursue only the antifoundationalist pragmatist view.

[121] The sweeping philosophical implications of rejecting traditional dichotomies between language and reality, and fact and value, may be difficult to imagine. They would infuse and transform our everyday discourse. The vocabulary of our conversation still presupposes certain categories and foundational principles that we are in the process of philosophically rejecting. *See, e.g.*, R. RORTY, *Pragmatism, Relativism, and Irrationalism*, in CONSEQUENCES OF PRAGMATISM 160 (1982).

[122] *See* H. PUTNAM, REASON, TRUTH AND HISTORY 139-41 (1981). What is at issue both for Putnam and for me is not any kind of directional causal chain, but rather only the coherence of values, facts, and discourse. *See, e.g.*, *id.* at 132-35, 201-03, 215. In other words, the argument is that the world is holistic and that these matters are all interdependent. Changes anywhere will lead to changes everywhere. Universal commodification presupposes, as well as leads to, an inferior conception of human flourishing. Putnam is interested in dissolving the fact/value dichotomy, whereas I am interested in establishing that rhetoric is implicated in our facts-and-values. I believe Putnam makes this point, using the term "conceptual schemes" rather than discourse or rhetoric, although he does not stress it. Hence, I believe my position to be substantially similar to his, although different in emphasis.

[123] *See id.* at 139-40. This term, referring to the aggregate level of satisfaction or pleasure, derives from the use of a hedonic calculus (like Bentham's) to judge the good.

> [W]hile they would not cause someone suffering for the sake of the greatest happiness of the greatest number if there were reasonable doubt that *in fact* the consequence of their action would *be* to bring about the greatest happiness of the greatest number, . . . in cases where one knows with certainty what the consequences of the actions would be, they would be willing to . . . torture small children or to condemn people for crimes which they did not commit if the result of these actions would be to increase the general satisfaction level in the long run . . . by any positive [increment], however small.[124]

Putnam says that the difference between us and the super-Benthamites is not merely a disagreement about values.[125] Our disagreement about values will entail disagreement about facts and descriptions of facts. For example, super-Benthamites would realize that sometimes the greatest happiness of the greatest number requires telling a lie; it would not count as dishonest in any pejorative sense to tell lies in order to maximize the general pleasure level. Nor would it be wrong to break promises that would not maximize pleasure if kept. The use of the term "honest" among super-Benthamites would be extremely different from our use of that same descriptive term.[126] Terms like "considerate," "good citizen," or "good person" would likewise be subject to different uses. The vocabulary for describing interpersonal situations would vary greatly between us and the super-Benthamites:

> Not only will they lack, or have altered beyond recognition, many of our descriptive resources, but they will very likely invent new jargon of their own (for example, exact terms for describing hedonic tones) that are unavailable to us. *The texture of the human world will begin to change*. In the course of time the super-Benthamites and we will end up living in different human worlds.
>
> In short, it will not be the case that we and the super-Benthamites "agree on the facts and disagree about values". In the case of almost all interpersonal situations, the description we give of the facts will be quite different from the description they give of the facts. Even if none of the statements they make about the situation are *false*, their description will not be one that we will count as adequate and perspicuous; and the description we give will not be one that they could count as adequate and perspicuous. In short, even if we put aside

---

[124] *Id.* at 140. Extreme utilitarians ("super-Benthamites") would in fact be led to argue for punishment of innocent people on consequentialist grounds. *Cf.* Smart, *Extreme and Restricted Utilitarianism*, 6 PHIL. Q. 344 (1956). But because we are *not* super-Benthamites, the idea of punishment of the innocent has instead been an embarrassment for utilitarianism, a problem it must solve. *See* Rawls, *Two Concepts of Rules*, 64 PHIL. REV. 3 (1955).

[125] H. PUTNAM, *supra* note 122, at 141.

[126] *See id.* at 140.

> our "disagreement about the values", we could not regard their total representation of the human world as fully rationally acceptable.[127]

Putnam concludes that the super-Benthamites' inability rightly to comprehend "the way the human world is" results from their "sick conception of human flourishing" — their inferior theory of the good for human beings.[128]

Putnam's parable is relevant to the conceptualization of rape as theft of a property right. It suggests that a particular conception of human flourishing is advanced by this pervasive use of market rhetoric. To think in terms of costs to the victim and her sympathizers versus benefits to the rapist is implicitly to assume that raping "benefits" rapists. Only an inferior conception of human flourishing would regard rape as benefiting the rapist. As a reason for criminalizing rape, Posner blandly says, "Supposing it to be true that some rapists would not get as much pleasure from consensual sex, it does not follow that there are no other avenues of satisfaction open to them."[129] The "pleasure" and "satisfaction" of maintaining one's bodily integrity is commensurate with the "pleasure" and "satisfaction" of someone who invades it. Thus, there could be circumstances in which the satisfactions or "value" to rapists would outweigh the costs or "disvalue" to victims.[130] In those situations rape would not be morally wrong and might instead be morally commendable.

Market rhetoric, if adopted by everyone, and in many contexts, would indeed transform the texture of the human world. This rhetoric leads us to view politics as just rent seeking, reproductive capacity as just a scarce good for which there is high demand, and the repugnance of slavery as just a cost. To accept these views is to accept the conception of human flourishing they imply, one that is inferior to the conception we can accept as properly ours.[131] An inferior conception of human flourishing disables us from conceptualizing the world rightly. Market rhetoric, the rhetoric of alienability of all "goods," is

---

[127] *Id.* at 141 (first emphasis added, second emphasis in original).

[128] *Id.* at 141. As Putnam reminds us, giving normative weight in this way to conceptions of human flourishing, *eudaemonia*, is a form of neo-Aristotelianism. *See id.* at 148. Other theorists who are pursuing normative conceptions of human nature or flourishing are Alasdair MacIntyre, *see* A. MACINTYRE, AFTER VIRTUE (2d ed. 1984), and Roberto Unger, *see* R. UNGER, PASSION (1984).

[129] Posner, *supra* note 50, at 1199.

[130] *See* R. POSNER, *supra* note 8, at 202 (quoted in note 109 above).

[131] In order to decide what conception of human flourishing is properly ours, all we can do is reflect on what we now know about human life and choose the best from among the conceptions available to us. *See infra* note 208. Thus, we should not accept a conception of human flourishing that excludes our understanding of politics as (also) community self-determination, excludes our understanding of reproductive capacity as essentially human and personal, and excludes our understanding that the pain of witnessing criminal acts and unjust institutions is not like the price of shoes or snowplows.

also the rhetoric of alienation of ourselves from what we can be as persons.[132]

One way to see how universal market rhetoric does violence to our conception of human flourishing is to consider its view of personhood. In our understanding of personhood we are committed to an ideal of individual uniqueness that does not cohere with the idea that each person's attributes are fungible, that they have a monetary equivalent, and that they can be traded off against those of other people. Universal market rhetoric transforms our world of concrete persons, whose uniqueness and individuality is expressed in specific personal attributes, into a world of disembodied, fungible, attribute-less entities possessing a wealth of alienable, severable "objects." This rhetoric reduces the conception of a person to an abstract, fungible unit with no individuating characteristics.[133]

Another way to see how universal market rhetoric does violence to our conception of human flourishing is to consider its view of freedom. Market rhetoric invites us to see the person as a self-interested maximizer in all respects. Freedom or autonomy, therefore, is seen as individual control over how to maximize one's overall gains. In the extreme, the ideal of freedom is achieved through buying and selling commodified objects in order to maximize monetizable wealth.[134] As we have seen, Marx argued with respect to those who produce and sell commodities that this is not freedom but fetishism; what and how much is salable is not autonomously determined.[135] Whether or not we agree with him, it is not satisfactory to think that marketing whatever one wishes defines freedom. Nor is it satisfactory to think that a theoretical license to acquire all objects one may desire defines freedom.[136]

To reject the slogan, "The word is not the thing," is not to deny that there is a difference between thought and action. To say "I wish you were dead" is not to kill you. Rather, rejecting the slogan is a way of understanding that the terms in which human life is conceived matter to human life. Understanding this, we must reject universal commodification, because to see the rhetoric of the market — the rhetoric of fungibility, alienability, and cost-benefit analysis — as the

---

[132] *See* M. Radin, The Rhetoric of Alienation (unpublished manuscript on file at the Harvard Law School Library).

[133] This conception of the person as an abstract, fungible unit makes us all interchangeable ghosts. It seems — as we shall see in Part IV — that something like this ghostly abstract conception of personhood has been popular in liberal political and moral theory for a long time. *Cf. supra* note 91. Nevertheless, I think we are not (and need not be) committed to this conception in the wholehearted way required by universal market rhetoric.

[134] Universal commodification is allied with a narrow view of negative freedom. *See infra* p. 1905. For further discussion of negative freedom, see Section C of Part IV below.

[135] *See supra* pp. 1872–73.

[136] *See infra* pp. 1904–06.

sole rhetoric of human affairs is to foster an inferior conception of human flourishing.

Of course, commodification-talk can still be one rhetoric among others. Yet once we accept the view that the terms of discourse affect our conception of human flourishing, we must decide when it is morally appropriate to think and speak in market rhetoric and when it is not. In this regard a problem arises for the liberal conception of free speech. A pluralist who holds that market rhetoric is inappropriate for describing or reasoning about some valued things might try to forestall the kind of transformation envisaged in Putnam's parable by curtailing some market rhetoric. If such a pluralist holds that market rhetoric should be prohibited in at least some cases, this view would conflict with the liberal ideal of neutrality with respect to choice of discourse.[137]

The liberal view of free speech becomes understandably difficult to maintain once one accepts the holistic view of rhetoric I have been discussing. In this view of rhetoric there can be no sharp distinction between speech and action or between content-based and content-neutral regulations. Just as there is no analytic divide between words and facts when rhetoric and reality are fused, so too there is no such divide between words and acts. The holistic view of rhetoric necessitates that we make choices in our discourse for constructing our human world of facts and values, like it or not. There cannot be a stance that is neutral on the good life for human beings. To deny the possibility of neutrality is of course to deny a central liberal tenet.[138]

---

[137] I am deliberately not taking the position that if one kind of discourse expresses and fosters an inferior conception of human flourishing, government censorship is necessarily justified. Unless we are convinced that pursuing the good is always the province of the government, it is certainly open to us to decide provisionally that certain discourses embody inferior conceptions of human flourishing without endorsing censorship of them. Moreover, even if pursuing the good is a collective duty, tolerance may be a good to be pursued. *Cf.* H. PUTNAM, *supra* note 122, at 49 (arguing that "respect for persons . . . requires that we accord them the right to choose a moral standpoint for themselves").

[138] Hence, a thoughtful pluralist who wishes to remain a liberal will encounter a deep tension. For example, Thomas Scanlon argues:

> [I]f what the partisans of pornography are entitled to (and what the restrictors are trying to deny them) is a fair opportunity to influence the sexual mores of the society, then it seems that they, like participants in political speech in the narrow sense, are entitled to at least a certain degree of access even to unwilling audiences.

Scanlon, *Freedom of Expression and Categories of Expression*, 40 U. PITT. L. REV. 519, 545–46 (1979). This conclusion might be avoided to some extent, Scanlon argues, if we could determine whether

> exposure . . . leads to changes in one's tastes and preferences through a process that is, like subliminal advertising, both outside of one's rational control and quite independent of the relevant grounds for preference, or whether, on the contrary, the exposure to such influences is in fact part of the best way to discover what one really has reason to prefer.

Id. at 547. Scanlon concludes, somewhat obscurely, that there may be grounds for protecting unwilling audiences against exposure to pornography involving violence, torture, or domination,

Although pluralism — accepting commodification-talk as one discourse among others — may be appropriate, we cannot be neutral about whether to conceive of any particular thing in market rhetoric.

IV. PLURALISM: THE LIBERAL HERITAGE

As this Part outlines, pluralism has been a prominent tenet of traditional liberalism.[139] Nevertheless, as this Part also attempts to show, liberal pluralism has borne within it the seeds of universal commodification; indeed, universal commodification is a more coherent liberal position than pluralism. Thus, I ultimately argue that pluralism should now be reconceived.

In order to understand why liberal pluralism should be reconceived, a review of its ideological heritage is necessary. Prominent principles in liberal pluralism include negative liberty, the person as abstract subject, and a conceptual notion of property. These princi-

---

but not against exposure to "mildly unconventional sexual attitudes or practices." *Id.* at 549. The reason given is that for the latter, "it is more plausible to say that discovering how one feels about such matters when accustomed to them is the best way of discovering what attitude towards them one has reason to hold." *Id.*

The argument that even if one likes violence, torture, and domination after being accustomed to them, this is not a *reason* for accepting them, *see id.* at 548, seems to me to be merely a disguised argument for rejecting an inferior conception of human flourishing, and the discourse (pornography) in which it is couched. There is nothing in the argument as explicitly given that would otherwise tell us when growing to like something is a reason to value it and allow it to change one's human world, and when not. Disguising the argument suppresses the dissonance caused by accepting a link between discourse and personhood while trying to remain a liberal.

In contrast, a significant strand of feminist thought holds the view that even willing audiences should not be allowed to shape and reinforce their view of human sexuality, and in particular of women's role in it, through pornography. *See* Brest & Vandenberg, *Politics, Feminism, and the Constitution: The Anti-Pornography Movement in Minneapolis*, 39 STAN. L. REV. 607 (1987); MacKinnon, *Pornography, Civil Rights, and Speech*, 20 HARV. C.R.-C.L. L. REV. 1 (1985). As this view does not try to retain the indicia of liberalism — in particular that one must be neutral on what constitutes the good life — it can be much more explicit about the reason for rejecting this kind of discourse: it reflects an inferior conception of human flourishing and one that is degrading to women. And this view sees no room for tolerance, *cf. supra* note 137, because it holds that a dominant inferior conception of human flourishing is simply preventing women from developing their own self-conception.

[139] That there should be a realm of inalienable "political" rights and a realm of alienable "property" rights seems fundamental to those who hold a traditional world view that divides up the social world into politics and markets. In the realm of politics there are familiar inalienable individual rights like life, liberty, and the pursuit of happiness; there are also autonomy, liberty, and equal respect for persons. In the realm of the market there are alienable property rights and free trade. Karl Polanyi thought this pluralist view to be necessary to the emergence of the full laissez-faire market society: "A self-regulating market demands nothing less than the institutional separation of society into an economic and political sphere. Such a dichotomy is, in effect, merely the restatement, from the point of view of society as a whole, of the existence of a self-regulating market." K. POLANYI, THE GREAT TRANSFORMATION 71 (1944).

ples are basic to the free market and its institutions, private property and free contract. Negative liberty and the subjectivity of personhood underlie convictions that inalienable things are internal to the person, and that inalienabilities are paternalistic. Conceptualism finds alienability to be inherent in the concept of property. These convictions make the case for liberal pluralism uneasy, always threatening to assimilate to universal commodification.

A. Inalienability and the Concept of Property

The legal infrastructure of capitalism — what is required for a functioning laissez-faire market system — includes not merely private property, but private property plus free contract.[140] In order for the exchange system to allocate resources, there must be both private entitlement to resources and permission to transfer entitlements at will to other private owners. Liberal theorists have expressed or reflected this necessity with conceptualist and separatist strategies. In a conceptualist strategy, both necessary characteristics can be incorporated either into the property theory, by claiming that free alienability is inherent in the concept of property, or into the contract theory, by claiming that private entitlement is inherent in the concept of freedom of contract. In a separatist strategy, property and contract split the capitalist indicia between them.[141]

Some writers, such as Hume, Kant, and Hegel, used a separatist strategy to justify private property and free contract.[142] In such discussions the justifications for entitlement and alienability, although separate, are interlocking parts of the same picture. Other writers, such as Mill, used a conceptualist strategy to assert that (market) alienability is inherent in the concept of private property.[143] This

---

[140] Kennedy and Michelman refer to the market regime as the "PPFC" regime. *See* Kennedy & Michelman, *Are Property and Contract Efficient?*, 8 HOFSTRA L. REV. 711, 750 (1980).

[141] The common law of restraints on alienation seems to reflect both the conceptualist and the separatist traditions. To strike down restraining conditions because they are said, without more, to be "repugnant to a fee," is merely to say that free alienability is implicit in the concept of the fee simple absolute. *See, e.g.*, Northwest Real Estate Co. v. Serio, 156 Md. 229, 234, 114 A. 245, 246 (1929) (holding as repugnant to the fee simple title a clause preventing the property from being sold or rented prior to a designated date without the consent of the grantor). When restraining conditions have been weighed to determine their "reasonableness," alienability is potentially a social variable separate from private ownership. *See, e.g.*, Wellenkamp v. Bank of Am., 21 Cal. 3d 943, 582 P.2d 970, 143 Cal. Rptr. 379, (1978) (holding that a due-on-sale clause in a deed of trust constituted an unreasonable restraint on alienation).

[142] *See infra* pp. 1891–98.

[143] *See infra* pp. 1889–91. One might think of positivist functionalism as a third strategy: property and contract have whatever characteristics we choose to give them in order to accomplish our goals. In this strategy, entitlement and free alienation are justified by more direct reference to the requirements of the free market. It stems from the work of Jeremy Bentham, *see* J. BENTHAM, THEORY OF LEGISLATION, PRINCIPLES OF THE CIVIL CODE pt. I, chs. 6–12 (R. Hildreth trans. 1840), and has many intellectual descendants, particularly among economists.

argument structure submerges the issue of alienability and makes justification of it seem less necessary: once property is justified, the task of justifying the market is done.[144]

The conceptualist strategy faces a problem: it cannot consistently admit any inalienabilities without denying that the objects of them are property. A conceptualist cannot admit that ownership is sometimes justified only when the object owned is beyond the reach of the market. John Stuart Mill's property theory illustrates the problem. Mill declared that "included in the idea of private property" is a right of each person "to the exclusive disposal of what he or she have produced by their own exertions, or received either by gift or by fair agreement without force or fraud, from those who produced it."[145] The right of property "includes . . . the freedom of acquiring by contract," because to prevent those who produce things from giving or exchanging them as they wish violates the producers' property rights.[146] Also "implied in property" is the right to whatever a producer can get for her products "in a fair market."[147] Taken together, these declarations establish that Mill's idea of property inherently requires contracts and markets. It would be a logical contradiction for him to postulate inalienable property.

Yet, in other passages, Mill argued for inalienabilities and for restraints on alienation. He argued that the laws of property "have made property of things which never ought to be property, and absolute property where only a qualified property ought to exist."[148] He

---

Positivist functionalism need not be thought of as a third strategy, however, because it, too, is either conceptualist or separatist. When it is conceptualist, private entitlement and free alienation are understood in what property *means* as well as in what property must be to facilitate the maximization of wealth, welfare, or utility through free exchange. *See infra* note 144. When it is separatist, private entitlement and free alienation are separately justified by their desired consequences. In this Part, I have not thought it necessary to elaborate how the theory of Bentham and his successors tends toward or facilitates universal commodification.

[144] Those who tend toward universal commodification often take this conceptual view. *See, e.g.*, R. EPSTEIN, *supra* note 48; Epstein, *supra* note 53; *cf.* Radin, *The Consequences of Conceptualism*, 41 U. MIAMI L. REV. 239 (1987) (arguing that Epstein's conceptualism has troublesome consequences for the justification of antidiscrimination rights). But conceptualism is not necessarily linked with the views of those who espouse commodification. Marx's "bourgeois property" is a similar conceptual notion meaning ownership plus free alienability, that is, commodification. *See, e.g.*, K. MARX, THE COMMUNIST MANIFESTO (P. Sweezey trans. 1964); Marx, *The German Ideology*, in THE MARX-ENGELS READER, *supra* note 81, at 186–93. Marx presumably would not have accepted any view that ownership might be justified if separated from market alienability, since he asserted that bourgeois property could not coexist with other kinds. *See* 3 K. MARX, *supra* note 85, at 505–19. To consider nonmarket "property" or inalienable "property" seems to be a contradiction both for Mill and for Marx..

[145] J.S. MILL, PRINCIPLES OF POLITICAL ECONOMY bk. II, ch. ii, at 218 (W. Ashley ed. 1909).

[146] *Id.* at 220.

[147] *Id.* at 221.

[148] *Id.* bk. III, ch. i, at 208.

thought that public offices, monopoly privileges, professional *brevets* — and human beings — should not be considered property.[149] In refusing to countenance certain things as property at all, Mill was able to avoid the contradiction that inalienable property poses for his conceptualism. Saying that some things are not property is one way for a liberal to be a pluralist. But Mill also thought that people could justifiably hold only "qualified," and not "absolute," property rights in land and presumably in other natural resources.[150] Here he could not avoid the contradiction. What he said about property in land implies some inalienability (and some curtailment of the right to exclude others), thus contradicting his general conceptual vision of property.[151]

One who thinks that some things can be "property," but not fully alienable, is a different kind of pluralist from one who holds that some things are not property at all. A conceptualist can be a pluralist by holding that some things are not property at all, but she cannot consistently be a pluralist by holding that some things that are property are not fully alienable. The logical contradiction invites a move from the latter kind of pluralism toward complete commodification: everything that is property must be fully alienable, because property is necessarily suitable for trade in a laissez-faire market. Mill's "qual-

---

[149] Political theorists who explain and justify capitalist private property must address the issue of human commodification. Their problem is to condemn slavery while justifying the sale of one's labor on the market model; that is, to distinguish worker commodification under slavery from the (alleged) worker commodification under (alleged) wage slavery. The sociologist Orlando Patterson argues that there is no intrinsic difference between "property" in the work of slaves, on the one hand, and in the work of employees or of divorced spouses with legally enforced support obligations, on the other. *See* O. PATTERSON, SLAVERY AND SOCIAL DEATH 21–27 (1982). Patterson's main point is that slavery is not tied to the notion of property in human beings; it exists under many kinds of social structures that do not include property. For those who wished to affirm the liberal market society and its pervasive property relations, while rejecting slavery, however, distinguishing between market property in human beings' labor and slave property in human beings' labor was crucial. For Hegel's difficulties with the problem of selling oneself into slavery, see p. 1894 below; for Mill's, see pp. 1902–03 below.

[150]

> The essential principle of property being to assure to all persons what they have produced by their labour and accumulated by their abstinence, this principle cannot apply to what is not the produce of labour, the raw material of the earth. . . . Whenever, in any country, the proprietor . . . ceases to be the improver, political economy has nothing to say in defense of landed property, as there established.

J.S. MILL, *supra* note 145, at 229–30, 231.

[151] For example, from the passage quoted in note 150 above, we might infer that land should not be owned by nonimprovers. The tension between the property rights Mill actually thought could be justified and what he said was required for something even to be property is also evident in Mill's views on bequest. He considered bequest "one of the attributes of property," *see* J.S. MILL, *supra* note 145, at 226–29, but was opposed to unlimited bequest. This seems to contradict both the unlimited right to dispose (for the testator) and the unlimited right to receive (for the devisee or legatee) that Mill asserted are inherent in the concept of property. *See* D. Thompson, Inheritance and Property Rights (unpublished manuscript on file at Harvard Law School Library).

ified" property rights would be qualified only if necessary to avoid externalities that would otherwise create market failure.[152] This position leads to the traditional liberal pluralist picture of a laissez-faire market domain walled off from a few exceptions that are completely removed from the market.[153] It approximates universal commodification if the list of things that cannot be property at all is short.

B. The Subject/Object Dichotomy: The Kantian Person Versus the Thing-In-Itself

Theorists who do not adopt the conceptual strategy avoid the problem of having to view all restraints on the laissez-faire market as incompatible with property. The separatist strategy practiced by Kant and Hegel justifies property and alienability (free contract) based on their connection with freedom and actualization of the person.[154] It asserts that only objects separate from the self are suitable for alien-

---

[152] Mill thought that land ownership yielded only "qualified" property because its importance and scarcity resulted in a duty of stewardship owed to society by its owners. *See* J.S. MILL, *supra* note 145, at 229–35. He also thought that landed property was less justifiable than property in things created by one's own faculties, *see supra* note 150, and hence gave rise to weaker or fewer rights. He might well have rejected the idea that the bad consequences of allowing land to be exclusively controlled by its owners should be characterized as external costs; he might well have rejected the idea that his qualifications were only exceptions in aid of market results.

[153] In the traditional liberal pluralist picture, the market domain is dominant, large, and is the rule; the nonmarket domain is subordinate, small, and is the exception. Inalienabilities are anomalies in social life, requiring special explanation and justification. This picture concedes much to the market and to commodification. In my view, a more satisfactory pluralism will relinquish conceptualism about property, and then it need not concede that there is any prima facie laissez-faire market realm on one side of a metaphorical wall. *See* Radin, *supra* note 42.

[154] A different kind of separatist strategy, which did not equate transfer of property with contract, was practiced by Hume in the first half of the eighteenth century. Hume's separate chapters on transfer of property and on the obligation of promises reflect his view that the exchanges of services and actions that are the subject of contract are part of the commerce of mankind although not commodities. *See* D. HUME, A TREATISE OF HUMAN NATURE bk. III, pt. II, § ii, at 484 (*Of the Origin of Justice and Property*), § iii, at 501 (*Of the Rules, which Determine Property*), § iv, at 514 (*Of the Transference of Property by Consent*), § v, at 516 (*Of the Obligation of Promises*) (L. Selby-Bigge 2d ed. 1978) (1740).

For Hume, private property is a convention for the stability of possessions. *See id.* § ii, at 484. But this convention — to respect others' possessions if they will respect your own — still leaves the "grand inconvenience" that possessions held by the recognized rules of acquisition are not likely to be in the hands of the most fitting person. *See id.* § iv, at 514. Because to "allow every man to seize by violence what he judges to be fit for him, wou'd [sic] destroy society," justice requires "some medium betwixt a rigid stability, and this changeable and uncertain adjustment." *Id.* This "medium" is the "translation of property by consent," and its necessity is "determined by a plain utility and interest." *Id.* at 514–15. Hume did not bring up any examples of circumstances in which "plain utility and interest" require that something be nontransferable, though reliance on utility and interest does leave this move open. Modern economic analysts are his successors: transaction costs explain or justify some inalienabilities in the world of presumptively alienable property. *See supra* pp. 1863–70.

ation. The problem confronting this separatist strategy is to distinguish things internal from things external to the person — the subject/object problem.

The subject/object dichotomy metaphysically divides the universe into opposed subjective and objective realms. Kantian personhood is the subject side of the dichotomy. Kantian persons are essentially abstract, fungible units with identical capacity for moral reason and no concrete individuating characteristics.[155] They are units of pure subjectivity acting in and upon the world of objects. Pluralism based on this conception of the person founders in trying to draw the distinction between persons and objects; once it does, it gravitates toward universal commodification. The Kantian conception of personhood makes us all interchangeable and thus facilitates liberal political equality. But by postulating such a world of fungible, subjective, autonomous units, it also facilitates conceiving of concrete personal attributes as commodified objects.

The difficulties caused by Kantian personhood can be seen by examining the German theory of property and contract, in which entitlement and alienability are separately justified, relying on the subject/object dichotomy. For Kant and Hegel, private property is necessary to realize or actualize the will of a person in order to achieve freedom.[156] The German theory posited that, in order to be free and well-developed selves, we must be able to alienate external things and we must not be able to alienate internal things.[157] For Hegel, alienability of property (both transfer and relinquishment) was not inherent in the concept of property, but rather followed from the premise that the presence of a person's will makes something property:

> The reason I can alienate my property is that it is mine only in so far as I put my will into it. Hence I may abandon . . . anything that I have or yield it to the will of another . . . provided always that the thing in question is a thing external by nature.[158]

It followed that whatever is mine but is *not* "a thing external by nature" will be inalienable (nonrelinquishable and nontransferable).

---

[155] *See* I. KANT, *General Introduction to the Metaphysic of Morals*, in THE PHILOSOPHY OF LAW 9, 31–32 (W. Hastie trans. 1887) (1797) ("A PERSON is a Subject who is capable of having actions *imputed* to him. Moral Personality is, therefore, nothing but the Freedom of a rational Being under Moral Laws"); G. HEGEL, PHILOSOPHY OF RIGHT § 41 (T. Knox trans. 1952) ("Personality is the first, still wholly abstract, delineation of the absolute and infinite will."); *id.* § 35 ("The universality of [the] consciously free will is abstract universality, the self-conscious but otherwise contentless and simple relation of itself to itself in its individuality, and from this point of view the subject is a person.").

[156] *See* I. KANT, *The Science of Right*, in THE PHILOSOPHY OF LAW, *supra* note 155, at 61, 61–67; G. HEGEL, *supra* note 155, §§ 41–71.

[157] *See infra* p. 1893.

[158] G. HEGEL, *supra* note 155, § 65.

Substantive characteristics of personality are not things external by nature and are hence inalienable.[159]

In order to apply this Hegelian argument to delineate inalienabilities, then, we need to draw clearly the distinction between things external by nature and substantive constitutive elements of personality. If something is external by nature, it must be propertizable and alienable so that persons can achieve freedom and proper self-development.[160] If something is a substantive characteristic of personality, it must be inalienable for the same reasons.[161] Thus, inalienability is either required or proscribed, and the decision turns on the distinction between external things and substantive characteristics.

By "things external by nature," Hegel meant objects in the environment that have (or can be thought to have) an existence independent of our will.[162] The initial gulf between the abstract will of the person and the world of unowned objects expresses the dichotomy between subject and object.[163] The gulf between subject and object

---

[159] What Hegel had in mind here are "personality as such," "universal freedom of will," "ethical life" (community morality), and "religion." *Id.* at § 66. (The term "ethical life" is a translation of Hegel's *Sittlichkeit*, the meaning of which might be better conveyed by "community morality.") These "goods, or rather substantive characteristics," constituting personality itself and the essence of self-consciousness are inalienable. *Id.* Under alienation of the personality itself, Hegel included slavery, serfdom, disqualification from holding property, encumbrances on property, "and so forth." *Id.* § 66R. Superstition, and ceding to someone else full power to direct one's actions or to prescribe duties of conscience or religious truth, "etc.," counted as forbidden alienation of intelligence and rationality, of morality and religion. *See id.*

[160] Kant and Hegel had similar views on propertization. *See* I. KANT, *supra* note 156, at 62 ("It is possible to have any external object of my Will as Mine. In other words, a Maxim to this effect — were it to become law — that any object on which the Will can be exerted must remain objectively in itself *without an owner* . . . is contrary to the Principle of Right." (emphasis in original)); G. HEGEL, *supra* note 155, § 44.

[161] Kant's position was similar to Hegel's. Kant had put the same subject/object problem in starker form:

> Man cannot dispose over himself because he is not a thing; he is not his own property; to say that he is would be self-contradictory; for in so far as he is a person he is a Subject in whom the ownership of things can be vested, and if he were his own property, he would be a thing over which he could have ownership. But a person cannot be a property and so cannot be a thing which can be owned, for it is impossible to be a person and a thing, the proprietor and the property.

I. KANT, LECTURES ON ETHICS 165 (L. Infield trans., J. Macmurray rev. ed. 1930). Kant's argument here was in the form of a contradiction, a form of argument that Hegel also used. *See infra* p. 1896. From this contradiction, which seems to rule out voluntary enslavement although Kant did not mention it here, Kant purported to deduce not only that sexual services cannot be marketed, but also that a person is not entitled to sell one of her teeth. *See* I. KANT, *supra*. If nothing else, this deduction can serve as a warning that the internal/external or subject/object distinction does not generate noncontroversial particular consequences.

[162] *See* G. HEGEL, *supra* note 155, § 42 ("What is immediately different from free mind is that which, both for mind and in itself, is the external pure and simple, a thing, something not free, not personal, without rights.")

[163] For Hegel, it is only an "initial" gulf because as *Geist* (mind or spirit) becomes actualized, the wills of persons will become actualized in objects, and objects will be enlivened by the wills

creates practical problems in deciding which items belong on which *side of the divide*; there are cases in which it does not seem intuitively obvious even to one who thinks the subject/object dichotomy itself is intuitively obvious.

Hegel's argument justifying wage labor (property in one's own capabilities for services) exhibits the difficulty. Faced with explaining why wage labor is justified while slavery is not, Hegel merely stated his desired conclusion that wage labor is "external" to personality:

> Single products of my particular physical and mental skill and of my power to act I can alienate to someone else and I can give him the use of my abilities for a restricted period, because, *on the strength of this restriction, my abilities acquire an external relation to the totality and universality of my being.* By alienating the whole of my time, as crystallized in my work, and everything I produced, I would be making into another's property the substance of my being, my universal activity and actuality, my personality.[164]

The argument that wage labor can become external because it is only part of one's creative capabilities and not the whole of them seems to be a non sequitur. Hegel did not claim that for something to be an internal, substantive characteristic of personality it had to be the whole of one's personal capacities. But even assuming the validity of the argument, it does not show that these capabilities are external by nature. Rather they seem to have gone from internal to external. Thus, Hegel's argument seems to contradict his own premise for alienability.[165] It is hard to see our work as belonging intrinsically to the object realm.

Hegel's flawed argument perhaps reflects an agenda: justifying the market. The market agenda, however, is not apparent on the face of his property theory. In fact, his theory might seem to be compatible

---

of persons. *See* G. HEGEL, *supra* note 155, § 44. For Kant, it is a permanent gulf; although persons (subjects) must control all things (objects) as property in order to be free, their character as subjects and objects is not thereby metamorphosed. *See* I. KANT, *supra* note 156, at 62 (quoted in note 160 above).

[164] G. HEGEL, *supra* note 155, § 67 (emphasis added).

[165] Hegel's argument for intellectual property exhibits a similar difficulty. Hegel thought that the method or medium of expression could externalize mental products and hence render them propertizable. *See id.* § 69. But this is not the same as saying they are a thing external by nature. In general, the concept of intellectual property is difficult to assimilate to the German model of justification of property. In fact, there had seemed to be no place at all for intangible property in Kant's theory of *possessio noumenon*. For Kant, there were only three propertizable external objects of the will: external corporeal things, another's free will in performance of a particular act, and certain status relationships. *See* I. KANT, *supra* note 156, at 64. Hegel followed Kant in relying on the external/internal (or subject/object) distinction, although he disavowed Kant's propertization of the status of wives and children. *See* G. HEGEL, *supra* note 155, §§ 75R, 161A, 175.

with a noncommodified view of society, because it is based upon embodiment of the will in objects and not upon trade, and because it blurs the line between subject and object.[166] For Hegel, objects may start out external, but they do not remain so: they become constitutive of personality. Indeed, the right to hold property is an inalienable attribute of personality.[167] Here the market agenda surfaces. Because personality is inalienable, one could argue that property might also be inalienable once personality is invested in it and constituted through it; but Hegel's argument is rather that any inalienability of property is itself a violation of inalienable personality rights.[168]

The underlying market agenda appears more clearly in Hegel's contract theory. Hegel said that "the concept" compels alienation of external objects *qua* property "in order that thereby my will may become objective to me as determinately existent."[169] The situation — contract — in which this compulsion of the concept is realized is the unity of different wills, and hence "the means whereby one identical will can persist within the absolute difference between independent property owners."[170] Other than the need to justify market relations, it is unclear why "reason" or "the concept" compel "gift, exchange, trade, etc.,"[171] just as it was unclear why wage labor is external to the person and why any inalienability of property rights would violate personality.[172]

---

[166] *See supra* note 163. For Hegel, the essence of property is just that it is necessary to embody the will and actualize personality:

> As *immediate* individuality, a person in making decisions is related to a world of nature directly confronting him, and thus the personality of the will stands over against this world as something subjective. For personality, however, as inherently infinite and universal, the restriction of being only subjective is a contradiction and a nullity. Personality is that which struggles to lift itself above this restriction and to give itself reality, or in other words to claim that external world as its own.

G. HEGEL, *supra* note 155, § 39 (emphasis in original). For a construal of Hegel's theory of property, see Radin, cited in note 115 above, at 971–78.

[167] *See* G. HEGEL, *supra* note 155, § 66R.

[168] Hegel argued that encumbrances on property (probably meaning what we would call restraints upon alienation, like conditions or entails) are themselves a disallowed alienation of substantive personality rights. *See supra* note 159. But note that Hegel thought that a landed aristocracy with entailed estates was most qualified to govern the properly developed state. *See* G. HEGEL, *supra* note 155, §§ 305–307; *cf. id.* § 180R ("In the higher sphere of the state, a right of primogeniture arises together with estates rigidly entailed; it arises, however, not arbitrarily but as the inevitable outcome of the Idea of the state."). In reconceiving pluralism I take the approach not taken by Hegel and argue that close connection to personhood can sometimes justify market-inalienability. *See infra* Part V.

[169] G. HEGEL, *supra* note 155, § 73. Something exists according to its concept (*Begriff*) when it is fully actualized in accord with mind or spirit (*Geist*).

[170] *Id.* § 74.

[171] *Id.* § 71R.

[172] If Hegel's property theory is a picture of the person's relationship with external objects, his contract theory is a picture of the person's relationship with other persons. Since it is a

Hegel also cast the argument against alienation of personhood as a "contradiction."[173] To alienate personhood is itself contrary to personhood, in that if I can relinquish my personhood, then no "I" remains to have done the relinquishing. If I treat "the infinite embodiment of self-consciousness" as something external and try to alienate it, Hegel argued, one of two things results: if I really possess these substantive attributes, they are not external and hence not alienated; if they are alienated, I did not possess them in the first place.[174] Hegel might have been trying to say that substantive personhood is simply not capable of objectification. The "contradiction" consists in supposing that one could give up that which, "so soon as I possess it, exists in essence as mine alone and not as something external."[175] If this interpretation is correct, then the contradiction poses the same subject/object problems as Hegel's general view of property and alienation: Why is it that personhood cannot be objectified while at the same time personhood requires objectification (in things)? Exactly what items are permanently "inside" the subject and incapable of objectification?

If the person/thing distinction is to be treated as a bright line that divides the commodifiable from the inalienable, we must know exactly which items are part of the person and which not. The person/thing distinction and its consequences seemed obvious to Kant and Hegel, but such is not the case for many modern philosophers.[176] One who

---

picture of two wills relating to each other in will-containing objects, no wonder Marx saw this kind of contract theory as fetishism. *See* 1 K. MARX, *supra* note 85, at 84–85.

Although Hegel argued that market exchange of property is required for proper self-development, he did not advocate universal commodification. Not only did he argue that certain things (namely, those belonging to substantive personality) were in principle not conceivable as property, but he also argued that family relationships and political relationships (the state) were not in principle conceivable as contract. *See* G. HEGEL, *supra* note 155, § 75R. In the progress of the ethical Idea from abstract to actual, the family and the state are higher spheres than the sphere of abstract right in which private property and free contract belong. *See id.* §§ 75R, 158–169, 261. The sphere of private right is the sphere of civil society (that is, the free market); the fully developed state is not merely an association of individual traders, but also (or rather) an organic entity, the embodiment of community morality, "the actuality of the ethical Idea" and "the actuality of concrete freedom." *See id.* §§ 183R, 257, 260. In these higher spheres, alienability and contract would be transcended by the advancing actualization of *Geist*. But it is open to dispute whether transcendence would mean that property and contract disappear, or that they continue to exist but with new significance. The latter is more likely the correct dialectical interpretation.

[173] G. HEGEL, *supra* note 155, § 66R.

[174] *See id.*

[175] *Id.*

[176] Nevertheless, the distinction and its consequences still seem obvious to some. For a recent discussion of inalienability that relies on an intuitive subject/object distinction, see Barnett, cited in note 4 above, at 195, in which the author states that "rights to possess, use, and control resources external to one's person are (generally) alienable, and . . . the right to possess, use, and control one's person is inalienable."

accepts the arguments of modern writers like Kuhn and Rorty[177] — and of Lukacs[178] before them — rejects the metaphysical bright line between what is inside us, in our minds, and some realm of things-in-themselves, a mind-independent reality outside of us. Without the bright line, arguments delineating the market realm on the basis of the subject/object distinction disintegrate. If the person/thing distinction is not a sharp divide, neither is inalienability/alienability. There will be a gray area between the two, and hence the outer contours of both personhood and inalienabilities based on personhood will remain contested.[179]

Pluralism's problematic reliance on the subject/object distinction can be submerged under universal commodification. Kantian notions of abstract personhood, particularly the conceptions of personhood as autonomous moral agency and persons as completely interchangeable units of subjective will, undergird liberal political ideals: equal treatment of persons as ends, not means; equality before the law; one person, one vote; and the rule of law.[180] These principles were the Enlightenment's great achievements, but a pull toward universal commodification seems to be the dark side of Kantian personhood. If the person is simply pure subjectivity empty of individuating characteristics and personal attributes, then these characteristics and attributes are readily conceived of as separate from the person and possessed by the person. From the view that attributes and characteristics are separate posessions, it is an easy step to conceptualize them as lying on the object side of the subject/object divide. This eliminates inalienabilities based on things internal to the person, because nothing is internal to the person considered as an abstract, subjective unit. Once individuating characteristics and personal attributes are conceptualized as possessions situated in the object realm,[181] it is another easy step to conceive of them as separable from the person through alienation.[182] Finally, once characteristics and attributes are seen as alienable objects, it is not difficult to see them as fungible and bearing implicit money value.[183] Kant no doubt would have abhorred this result;[184] nevertheless, universal commodification seems to be facili-

---

[177] *See* T. KUHN, *supra* note 120, at 6 (arguing that the adoption of scientific theories has both reflected and transformed "the world within which scientific work was done"); R. RORTY, *supra* note 120.

[178] *See* G. LUKÁCS, *supra* note 88.

[179] Traditional pluralism posits a bright line or wall between the market and nonmarket realms that is like the bright line dividing persons and things. *See supra* note 153.

[180] *See, e.g.*, Kuflik, *The Inalienability of Autonomy*, 130 PHIL. & PUB. AFF. 271, 296–98 (1984).

[181] *See supra* p. 1885.

[182] In a separatist strategy based upon the internal/external distinction, alienability is connected to situation in the external object realm. *See supra* pp. 1892–93.

[183] *See supra* pp. 1880–81.

[184] *See supra* note 161.

tated — though not entailed — by his definition of the person. The subject/object problem pulls pluralism toward universal commodification because there is no obvious stopping place short of that.[185]

C. Negative Liberty

Two theories about freedom are central to the ideological framework in which we view inalienability: the notion that freedom means negative liberty,[186] and the notion that (negative) liberty is identical with, or necessarily connected to, free alienability of everything in markets. The conception of freedom as negative liberty gives rise to the view that all inalienabilities are paternalistic limitations on freedom. The idea that liberty consists in alienability of everything in markets clashes with substantive requirements of personhood, making it difficult, for example, to argue against human commodification. In general, the commitment to negative liberty, like the commitment to the Kantian structure of persons versus objects, has caused confusion in liberal pluralism and has exerted a pull toward universal commodification.

Inalienabilities are often said to be paternalistic.[187] Paternalism usually means to substitute the judgment of a third party or the government for that of a person on the ground that to do so is in that

---

[185] Of course, the subject/object problem could pull in the opposite direction as well. If the bright line ceases to be intuitively obvious, yet one is not yet ready to give up the subject/object dichotomy entirely, one could retreat to a position placing (almost) everything in the subject realm as well as to a position placing (almost) everything in the object realm. Surrounding circumstances have so far rendered universal commodification the more popular philosophical refuge from the incomplete breakdown of the subject/object dichotomy.

[186] "Negative liberty" means roughly the freedom of the individual to be let alone to do whatever she chooses as long as others are not harmed. *See* I. BERLIN, *Two Concepts of Liberty* in FOUR ESSAYS ON LIBERTY 122 (1969) ("Political liberty in this sense is simply the area within which a man can act unobstructed by others."); *see also* Skinner, *The Idea of Negative Liberty: Philosophical and Historical Perspectives*, in PHILOSOPHY IN HISTORY 193, 197 (R. Rorty, J. Schneewind & Q. Skinner eds. 1984) (defining negative liberty as "the mere nonobstruction of individual agents in the pursuit of their chosen ends"). The positive/negative distinction, which Isaiah Berlin says is traditional, was used by Kant, who referred to the kind of arbitrary freedom of the will that we perceive in the phenomenal realm as negative. Positive freedom, in the noumenal realm, was for Kant identical to action necessitated by universal reason in conformity with moral law. *See* I. KANT, *supra* note 155, at 36. For a criticism of the concept of negative liberty, see Taylor, *What's Wrong with Negative Liberty*, in THE IDEA OF FREEDOM 175 (A. Ryan ed. 1979). For a useful taxonomy of various positive and negative conceptions of freedom, as well as a sophisticated defense of a modified negative view, see R. FLATHMAN, THE PHILOSOPHY AND POLITICS OF FREEDOM (1987).

[187] Calabresi and Melamed's discussion of inalienability rules illustrates a typical use of the notion of paternalism. *See supra* note 59. Another illustration is Anthony Kronman's treatment of restrictions on alienation as a form of paternalism. *See* Kronman, *Paternalism and the Law of Contracts*, 92 YALE L.J. 763 (1983).

person's best interests.[188] For advocates of negative liberty, to substitute someone else's choice for my own is a naked infringement of my liberty.[189] Freedom means doing (or not doing) whatever I as an individual prefer at the moment, as long as I am not harming other people.[190] To think of inalienability as paternalism assumes that freedom is negative liberty — that people would choose to alienate certain things if they could, but are restrained from doing so by moral or legal rules saying, in effect, that they are mistaken about what is good for them.

To say that inalienabilities involve a loss of freedom also assumes that alienation itself is an act of freedom, or is freedom-enhancing.[191] Someone who holds this view and conceives of alienation as sale through free contract is deeply committed to commodification as expressive of — perhaps necessary for — human freedom. Insofar as theories of negative freedom are allied to universal commodification, so are traditional discussions of inalienability in terms of paternalism. If we reject the notion that freedom means negative liberty, and the notion that liberty and alienation in markets are identical or necessarily connected, then inalienability will cease to seem inherently paternalistic. If we adopt a positive view of liberty that includes proper self-development as necessary for freedom, then inalienabilities needed to foster that development will be seen as freedom-enhancing rather than as impositions of unwanted restraints on our desires to transact in markets.

---

[188] As Duncan Kennedy points out, paternalism involves false consciousness. *See* Kennedy, *Distributive and Paternalist Motives in Contract and Tort Law, with Special Reference to Compulsory Terms and Unequal Bargaining Power*, 41 MD. L. REV. 563, 626–29, 631–49 (1982). The paternalist asserts that the actor has made a mistake about what is best for her and that a third party or the government is in a state of true consciousness and can therefore override her choice. Although Kennedy seeks to rescue it for his own purposes, the term "paternalism" has largely been used as a pejorative by advocates of negative liberty.

[189] Paternalism is particularly anathema to libertarians who are also moral subjectivists. They hold that a person's subjective preferences define her interests and, therefore, that it is nonsensical to claim that anyone else knows better than she does what is good for her. For these libertarians, among whose number are many practitioners of law and economics, the notion of false consciousness is simply incoherent.

Donald Regan argues that paternalism might be justified in some cases by converting the notion of freedom into a teleological principle (maximizing freedom), and that this is still a notion of negative freedom. *See* Regan, *Paternalism, Freedom, Identity, and Commitment*, in PATERNALISM 113–17 (R. Sartorius ed. 1983). Regan also proposes a form of justification based upon avoiding harm to someone's later self. This form of justification implicitly relies upon a notion of fostering personhood or self-development that may be inconsistent with negative liberty. An "Argument from Personal Integrity" close to Regan's is embraced in J. KLEINIG, PATERNALISM 67–73 (1984), and a similar argument is made by Kronman, cited in note 187 above, at 786–97.

[190] This is a rough restatement of the characteristic idea of negative liberty. *See supra* note 186.

[191] Both Mill and Hegel, at least in certain contexts, thought that alienation of objects is an act partly constitutive of individual freedom. *See supra* p. 1889, 1892. *But cf. infra* note 204.

Joel Feinberg's discussion of the inalienable right to life[192] illustrates the traditional link between inalienability and paternalism, as well as the tension caused by the clash between negative liberty and substantive requirements of personhood. Feinberg distinguishes three conceptions of the inalienable right to life, which he calls "the paternalist," "the founding fathers," and "the extreme antipaternalist."[193] In the view he calls paternalist, to say that the right to life is inalienable means that it is a nonrelinquishable mandatory right, one that ought to be exercised, like the right to education.[194] In contrast, the view that Feinberg attributes to the founding fathers holds that the inalienable right to life is a nonrelinquishable discretionary right.[195] It is discretionary because the individual may choose whether to exercise it.[196] For example, the right to own property is a discretionary right because I may choose to own nothing; it is a nonrelinquishable discretionary right because I cannot morally or legally renounce the right to own property even if I choose not to own any.[197] Feinberg concludes that the nonrelinquishable right to life is discretionary, not mandatory:

> [W]e have a right, within the boundaries of our own autonomy, to live or die, as we choose . . . . [T]he basic right underlying each is the right to be one's own master, to dispose of one's own lot as one chooses, subject of course to the limits imposed by the like rights of others. . . . In exercising my own choice in these matters, I am not renouncing, abjuring, forswearing, resigning, or relinquishing my right to life; quite the contrary, I am *acting* on that right by exercising it one way or the other.[198]

This passage suggests that the right to life is discretionary because it is parasitic on negative liberty. But Feinberg does not say whether the underlying right to be one's own master is mandatory or discretionary. The omission points to an apparent contradiction in the argument, a contradiction that stems from a commitment to negative liberty.[199] If the discretionary right to life is nonrelinquishable, as

---

[192] *See* Feinberg, *Voluntary Euthanasia and the Inalienable Right to Life*, 7 PHIL. & PUB. AFF. 93 (1978).

[193] *See id.* at 120–23.

[194] *See id.* at 120–21.

[195] Feinberg understands "inalienable" to mean prohibition of voluntary relinquishment. *See id.* at 112 ("An *inalienable right* is one that a person cannot give away or dispense with through his own deliberate choice." (emphasis in original)). In order to avoid unnecessary confusion, the discussion in the text substitutes "nonrelinquishable" for Feinberg's use of "inalienable."

[196] *See id.* at 121.

[197] *See id.* at 115–16.

[198] *Id.* at 121.

[199] Whether Feinberg is committed to negative liberty is unstated in the article under consideration, although that seems fairly inferable from his declaration of "doubts about the theory of inalienable rights in any case," *id.* at 94, and his characterization of mandatory rights as

Feinberg claims is the founding fathers' view, then we can infer that the "basic right" to *have* discretion — liberty — must be mandatory: one cannot choose not to be one's own master, not to dispose of one's lot as one chooses. But to attribute this mandatory conception of liberty to the founding fathers would apparently be to attribute to them a form of positive liberty, a view that people can be required to be free.[200] Hence, Feinberg attributes to the founding fathers a discretionary, not mandatory, view of the right to liberty. But if the right to liberty is indeed discretionary, then it seems I could choose *not* to be my own master, *not* to dispose of my lot as I choose, just as I could choose not to own property. And if I could choose that, I could choose not to have any of the other parasitic nonrelinquishable rights, like the right to life. The right to life would then be relinquishable.

This contradiction shows why a commitment to negative liberty pulls liberal pluralists toward universal commodification. The commitment to negative liberty usually attributed to the founding fathers forces those who hold it to choose between submerging a contradiction and moving toward conceiving of everything as relinquishable. If the intellectual descendants of the founding fathers want to maintain a nonrelinquishable discretionary right to life, they must adopt a mandatory right to liberty: we are not free not to be free. But adopting a mandatory right moves toward positive liberty, undermining the negative view that generates the nonrelinquishable, but discretionary, right to life. Holding firm to the view that liberty means negative liberty leads to a view that everything, including one's life, is relinquishable.

In this latter view, that of Feinberg's "extreme antipaternalist," the fully informed autonomous individual could sell herself into slavery or sell her right to life. Thus, the antipaternalist is a universal commodifier. This appears to be a more cogent view, once we grant that rights to life and property are parasitic upon an inalienable, but nonmandatory, right to negative liberty.

Might one hold fast to negative liberty and — contrary to the argument I have just given — still claim we are not free not to be free? This difficulty is the root of the tension between pluralism and negative liberty, and of the consequent pressure to give up pluralism.

"smug paternalism," *id.* at 122, and "offensively demeaning," *id.* at 106. A commitment to negative liberty is clear in J. FEINBERG, HARM TO SELF 62–66 (1986), in which Feinberg distinguishes among autonomy, liberty, and freedom, and defines both liberty and freedom in terms of absence of constraint.

[200] Contrary to the adherents of negative liberty, perhaps it is indeed plausible to attribute some form of positive liberty to the founding fathers. *See* Michelman, *The Supreme Court, 1985 Term — Foreword: Traces of Self-Government*, 100 HARV. L. REV. 4, 25–31, 47–55, 74 (1986).

Mill's well-known attempt to argue against freedom to sell oneself into slavery directly poses this difficulty:

> [B]y selling himself for a slave, [a person] abdicates his liberty; he foregoes any future use of it beyond that single act. He therefore defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself . . . . The principle of freedom cannot require that he should be free not to be free. It is not freedom, to be allowed to alienate his freedom.[201]

The argument is obscure. It is hard to see why Mill thought it obvious that the principle of negative freedom could not require the "freedom not to be free;" only positive freedom clearly holds that a person *must* be free.[202] In general, what in Mill's view is the connection between free alienation and freedom? (Why is alienation of freedom "not freedom"?) Most commentators have viewed Mill's argument against selling off one's freedom as a lapse into paternalism.[203]

Neither in his conception of freedom nor in his conception of alienability does Mill appear to explain why human beings are noncommodifiable. One could understand him to imply that there is an unstated divide between the realm of the market (free trade) and the realm of politics (liberty).[204] People must be free in order for a free political order to exist; they cannot be free without such a political order; hence, in the nonmarket realm they cannot, without contradic

---

[201] J.S. MILL, *On Liberty*, in THREE ESSAYS 126 (1975).

[202] *See supra* note 186.

[203] *See, e.g.*, J. FEINBERG, *supra* note 199, at 75–79.

[204] Because Mill insisted so strongly on the inherent alienability of property, *see supra* p. 1889, it is interesting that he elsewhere declared that "the principle of individual liberty is not involved in the doctrine of Free Trade." J.S. MILL, *On Liberty*, *supra* note 201, at 117. Hence, for Mill (unlike modern proponents of negative liberty), individual liberty is not involved in most government regulation of trade in commodities. Most trade restrictions, including restrictions on production, are wrong for Mill not because they violate the producers' liberty but because "they do not really produce the results which it is desired to produce by them." *Id.* They are wrong for utilitarian, not libertarian, reasons. On the other hand, prohibitions, "where the object of the interference is to make it impossible or difficult to obtain a particular commodity," do violate individual liberty, but that of the buyer, not the seller. *Id.* A modern version of this argument is to be found in Baker, *Counting Preferences in Collective Choice Situations*, 25 UCLA L. REV. 381 (1978), in which the author defends a distinction between regulation and prohibition that parallels Mill's, and in Baker, cited in note 87 above.

This argument seems to make the existence of a liberty interest depend on the motive with which the restraints are enacted. It also seems to imply that freedom is implicated in acquisition of goods, but not in disposition of them. Recall that in Mill's property discussion, both the right to dispose and the right to acquire were stressed as inherent in the idea of property. *See supra* p. 1889. If freedom is implicated in acquisition rather than disposition of goods, the idea that alienability and negative freedom are identical or necessarily linked is undermined. If freedom is primarily implicated in acquisition, then perhaps we should ask, as Mill did not, whether the would-be slave-owner's freedom instead of (or as well as) the would-be slave's is being violated by prohibiting slavery.

tion, be free not to be free. This reconstruction makes Mill a pluralist, as indeed he apparently wished to be; but the reading is not very true to Mill in the way it relinquishes negative liberty.

Again, one way to avoid Mill's problem is to espouse universal commodification. The universal commodifier can hold on to negative liberty and avoid Mill's problem — espousing negative liberty while eschewing voluntary enslavement — because under universal commodification freedom itself is seen as monetizable and alienable. Those who tend toward universal commodification may indeed endorse voluntary enslavement.[205] Those who declare human beings noncommodifiable must do so on the ground of postulated market failure (for example, transaction costs).[206]

We can now see why liberal pluralism should be reconceived. If we are to avoid the tendency toward universal commodification inherent in liberal pluralism, we must cease thinking that market alienability is inherent in the concept of property, and we must modify pluralism's commitments to negative liberty and Kantian personhood. In doing so, we must find a satisfactory way of deciding what market-inalienabilities are justified by the need to protect and foster personhood, and a way of understanding why these inalienabilities seem to us to be freedom enhancing.

V. TOWARD AN EVOLUTIONARY PLURALISM

In this Part, I develop a pluralist view that differs in significant respects from liberal pluralism. My central hypothesis is that market-inalienability is grounded in noncommodification of things important to personhood. In an ideal world markets would not necessarily be abolished, but market-inalienability would protect all things important to personhood. But we do not live in an ideal world. In the nonideal world we do live in, market-inalienability must be judged against a background of unequal power. In that world it may sometimes be better to commodify incompletely than not to commodify at all. Market-inalienability may be ideally justified in light of an appropriate conception of human flourishing, and yet sometimes be unjustifiable because of our nonideal circumstances.

A. Noncommodification and the Ideal of Human Flourishing

1. Rethinking Personhood: Freedom, Identity, Contextuality. — Because of the ideological heritage of the subject/object dichotomy,

---

[205] Robert Nozick takes the extreme view: a "free system" will allow an individual "to sell himself into slavery." R. NOZICK, *supra* note 48, at 331.

[206] *See* Demsetz, *supra* note 60, at 348–49; *supra* pp. 1865–66. For the reasons discussed in Section B of Part II above, this is deeply unsatisfactory from the point of view of personhood.

we tend to view things internal to the person as inalienable and things external as freely alienable. Because of the ideological heritage of negative liberty, we also tend to think of inalienabilities as paternalistic. A better view of personhood, one that does not conceive of the self as pure subjectivity standing wholly separate from an environment of pure objectivity, should enable us to discard both the notion that inalienabilities relate only to things wholly subjective or internal and the notion that inalienabilities are paternalistic.

In searching for such a better view, it is useful to single out three main, overlapping aspects of personhood: freedom, identity, and contextuality. The freedom aspect of personhood focuses on will, or the power to choose for oneself. In order to be autonomous individuals, we must at least be able to act for ourselves through free will in relation to the environment of things and other people.[207] The identity aspect of personhood focuses on the integrity and continuity of the self required for individuation. In order to have a unique individual identity, we must have selves that are integrated and continuous over time. The contextuality aspect of personhood focuses on the necessity of self-constitution in relation to the environment of things and other people. In order to be differentiated human persons, unique individuals, we must have relationships with the social and natural world.

A better view of personhood — a conception of human flourishing that is superior to the one implied by universal commodification — should present more satisfactory views of personhood in each of these three aspects. I am not seeking here to elaborate a complete view of personhood. Rather, I focus primarily on a certain view of contextuality and its consequences: the view that connections between the person and her environment are integral to personhood. I also suggest that to the extent we have already accepted certain views of freedom, identity, and contextuality, we are committed to a view of personhood that rejects universal commodification.[208]

---

[207] Because my purpose here is to sketch a rough, common understanding rather than to be philosophically precise, I do not attempt to elaborate a complete view of freedom, nor to characterize a distinction between freedom and autonomy, nor to define free will.

[208] The evolutionary pluralism that I recommend might also be called pragmatic pluralism, because it endorses a nonideal, immanent, and relatively particularist analysis. As with any pragmatic understanding, I cannot prove by any abstract principle that we implicitly accept certain views of personhood; I can only appeal to our most considered judgment and deepest sensitivity. By "our" deepest sensitivity I mean the sensitivity of all those who are engaged in this conversation, who find this conversation perspicuous, who think it is about the right things — even if my present views of them seem maddeningly wrong-headed.

Indeed, on a pragmatic understanding there is no one best (in the sense of ultimate and final) view of personhood, at least that we can currently understand; there are only views that we can come to recognize as better than those we have previously held. Thus, I can only seek to have us recognize a better view of personhood, not to formulate the best one. Some pragmatists think that even in the long run there is no "really" best view, whereas others think that there is in principle a "real" limiting concept, a "best" view that evolution progresses

Universal commodification conceives of freedom as negative liberty, indeed as negative liberty in a narrow sense, construing freedom as the ability to trade everything in free markets. In this view, freedom is the ability to use the will to manipulate objects in order to yield the greatest monetizable value. Although negative liberty has had difficulty with the hypothetical problem of free choice to enslave oneself,[209] even negative liberty can reject the general notion of commodification of persons: the person cannot be an entity exercising free will if it is a manipulable object of monetizable value.[210]

A more positive meaning of freedom starts to emerge if one accepts the contextuality aspect of personhood. Contextuality means that physical and social contexts are integral to personal individuation, to self-development. Even under the narrowest conception of negative liberty, we would have to bring about the social environment that makes trade possible in order to become the persons whose freedom consists in unfettered trades of commodified objects. Under a broader negative view that conceives of freedom as the ability to make oneself what one will, contextuality implies that self-development in accordance with one's own will requires one to will certain interactions with the physical and social context because context can be integral to self-development. The relationship between personhood and context requires a positive commitment to act so as to create and maintain particular contexts of environment and community. Recognition of the need for such a commitment turns toward a positive view of freedom, in which the self-development of the individual is linked to pursuit of proper social development, and in which proper self-development, as a requirement of personhood, could in principle sometimes take precedence over one's momentary desires or preferences.

Universal commodification undermines personal identity by conceiving of personal attributes, relationships, and philosophical and moral commitments as monetizable and alienable from the self. A better view of personhood should understand many kinds of particu-

---

toward. Richard Rorty is an example of the more skeptical kind of pragmatist, *see* R. RORTY, *supra* note 120, and Hilary Putnam is an example of the more realist kind, *see* H. PUTNAM, *supra* note 122. John Stick has recently placed the skeptical strain in the tradition of John Dewey and the realist strain in the tradition of Charles S. Peirce. *See* Stick, *Can Nihilism Be Pragmatic?*, 100 HARV. L. REV. 332, 341 n.27 (1986). In this Article, I remain uncommitted to either of these two strains. I do not assert either that there is potentially or in the long run one best concept of human flourishing, or that there is not. Although this is not the place to try to develop a complete metaethics, the problem has always struck me as one that defines the limits of our imagination by posing two possibilities, neither of which we can comfortably affirm. Questions about time and space are analogous: it seems mind-boggling either to affirm that time has been going on forever, or that there was once a first moment; it seems equally mind-boggling either to affirm that space goes on and on, or that it ends somewhere.

[209] *See supra* pp. 1902–03.

[210] *Cf. supra* note 161.

lars — one's politics, work, religion, family, love, sexuality, friendships, altruism, experiences, wisdom, moral commitments, character, and personal attributes — as integral to the self. To understand any of these as monetizable or completely detachable from the person — to think, for example, that the value of one person's moral commitments is commensurate or fungible with those of another, or that the "same" person remains when her moral commitments are subtracted — is to do violence to our deepest understanding of what it is to be human.[211]

To affirm that work, politics, or character is integral to the person is not to say that persons cease to be persons when they dissociate themselves from their jobs, political engagements, or personal attributes. Indeed, the ability to dissociate oneself from one's particular context seems integral to personhood.[212] But if we must recognize the importance of the ability to detach oneself, we must recognize as well that interaction with physical and social contexts is also integral to personhood. One's surroundings — both people and things — can become part of who one is, of the self. From our understanding that attributes and things can be integral to personhood, which stems mainly from our understanding of identity and contextuality, and from our rejection of the idea of commodification of the person, which stems mainly from our understanding of freedom, it follows that those attributes and things identified with the person cannot be treated as completely commodified. Hence, market-inalienability may attach to things that are personal.

---

[211] The Kantian person as a fungible unit of free will also is not a person as we know one. *See supra* note 155 and accompanying text. The person of Rawls's original position is also a Kantian unit devoid of most characteristics that situate us and make us human persons. *See* J. RAWLS, *supra* note 120. Rawls, of course, understands that real persons are different from the abstraction, but maintains that the Kantian abstraction is nevertheless appropriate for structuring liberal political and legal institutions. *See* Rawls, *Kantian Constructivism in Moral Theory*, 77 J. PHIL. 515, 533–35 (1980); Rawls, *Justice as Fairness: Political Not Metaphysical*, 14 PHIL. & PUB. AFF. 223, 232–34 (1985); *cf. supra* p. 1897. Rejecting liberal theories of political equality and the rule of law involves rejecting Kantian abstract personhood. *See, e.g.*, M. SANDEL, LIBERALISM AND THE LIMITS OF JUSTICE (1982); *cf.* Baker, *Sandel on Rawls*, 134 U. PA. L. REV. 895 (1985).

[212] Roberto Unger may be right in saying that continual transcendence of contexts is one mark of personhood. *See* R. UNGER, *supra* note 128. For a critique of Unger's notion of human nature as self-transformability, see R. Garet, The Critique of Human Nature (unpublished manuscript on file at Harvard Law School Library). To Unger's vision we must add that, paradoxically, a continuing commitment to contexts is also a mark of personhood. To be a person one must stay the same, but one must also change and develop. One cannot change everything all the time and be a person, nor can one change nothing ever and be a person. But we can recognize that persons must change without thinking of them as subjects completely separate from the "outside" world. Understanding that persons change themselves is not the same thing as thinking that the person is the subjective ghost that remains after everything that makes her what she is — although only for the present — is detached and thought of as an object separate from her.

2. Protecting Personhood: Noncommodification of Personal Rights, Attributes, and Things. — In my discussion of possible sources of dissatisfaction with thinking of rape in market terms,[213] I suggested that we should not view personal things as fungible commodities. We are now in a better position to understand how conceiving of personal things as commodities does violence to personhood, and to explore the problem of knowing what things are personal.

To conceive of something personal as fungible assumes that the person and the attribute, right, or thing, are separate. This view imposes the subject/object dichotomy to create two kinds of alienation. If the discourse of fungibility is partially made one's own, it creates disorientation of the self that experiences the distortion of its own personhood. For example, workers who internalize market rhetoric conceive of their own labor as a commodity separate from themselves as persons; they dissociate their daily life from their own self-conception. To the extent the discourse is not internalized, it creates alienation between those who use the discourse and those whose personhood they wrong in doing so. For example, workers who do not conceive of their labor as a commodity are alienated from others who do,[214] because, in the workers' view, people who conceive of their labor as a commodity fail to see them as whole persons.

To conceive of something personal as fungible also assumes that persons cannot freely give of themselves to others. At best they can bestow commodities. At worst — in universal commodification — the gift is conceived of as a bargain.[215] Conceiving of gifts as bargains not only conceives of what is personal as fungible, it also endorses the picture of persons as profit-maximizers. A better view of personhood should conceive of gifts not as disguised sales, but rather as expressions of the interrelationships between the self and others. To relinquish something to someone else by gift is to give of yourself. Such a gift takes place within a personal relationship with the recipient, or else it creates one.[216] Commodification stresses separateness both between ourselves and our things and between ourselves and other people. To postulate personal interrelationship and communion requires us to postulate people who can yield personal things to other

---

[213] *See supra* pp. 1880–81.

[214] *See* M. Radin, *supra* note 132.

[215] The universal commodifier would conceive of a gift as an exchange by assuming that giving you something that I value yields me monetizable value in return, or by assuming that I am doing it so that you will treat me with favor, and this favorable treatment yields monetizable value to me.

[216] *See* L. HYDE, *supra* note 101, at 56. "It is the cardinal difference between gift and commodity exchange that a gift establishes a feeling-bond between two people, while the sale of a commodity leaves no necessary connection." Gifts are similarly characterized by John Noonan as "given in a context created by personal relations to convey a personal feeling." J. NOONAN, BRIBES 695 (1984) (distinguishing among gifts, tips, contributions, and bribes).

people and not have them instantly become fungible. Seen this way, gifts diminish separateness. This is why (to take an obvious example) people say that sex bought and paid for is not the same "thing" as sex freely shared.[217] Commodified sex leaves the parties as separate individuals and perhaps reinforces their separateness; they only engage in it if each individual considers it worthwhile. Noncommodified sex ideally diminishes separateness; it is conceived of as a union because it is ideally a sharing of selves.

Not everything with which someone may subjectively identify herself should be treated legally or morally as personal.[218] Otherwise the category of personal things might collapse into "consumer surplus": anything to which someone attached high subjective value would be personal. The question whether something is personal has a normative aspect: whether identifying oneself with something — constituting oneself in connection with that thing — is justifiable. What makes identifying oneself with something justifiable, in turn, is an appropriate connection to our conception of human flourishing. More specifically, such relationships are justified if they can form part of an appropriate understanding of freedom, identity, and contextuality. A proper understanding of contextuality, for example, must recognize that, although personhood is fostered by relations with people and things, it is possible to be involved too much, or in the wrong way, or with the wrong things.

To identify something as personal, it is not enough to observe that many people seem to identify with some particular kind of thing, because we may judge such identification to be bad for people. An example of a justifiable kind of relationship is people's involvement with their homes. This relationship permits self-constitution within a stable environment. An example of an unjustifiable kind of relationship is the involvement of the robber baron with an empire of "property for power."[219] The latter is unjustified because it ties into a conception of the person we can recognize as inferior: the person as self-interested maximizer of manipulative power.[220]

There is no algorithm or abstract formula to tell us which items are (justifiably) personal. A moral judgment is required in each case. We have seen that Hegel's answer to a similar problem was to fall

---

[217] *See, e.g.*, F. HIRSCH, SOCIAL LIMITS TO GROWTH app. at 95–101 (1976).

[218] Those who subjectively identify with things not properly personal might be said to be alienated; improper object-relations keep them from being integrated persons according to the conception of human flourishing we accept. *See supra* pp. 1871–72, 1907.

[219] The distinction between "property for use" and "property for power" appears in Hobhouse, *The Historical Evolution of Property, in Fact and in Idea*, in PROPERTY: ITS DUTIES AND RIGHTS 3, 9–11, 23 (2d ed. 1922).

[220] The normative element in identifying personal things is discussed in a little more detail in Radin, *Residential Rent Control*, 15 PHIL. & PUB. AFF. 350 (1986), and in Radin, cited in note 115 above.

back on the intuition that some things are "external" and some are "internal."[221] This answer is unsatisfactory because the categories "external" and "internal" should be the conclusion of a moral evaluation and cannot be taken as obvious premises forming its basis. First we must judge whether persons can still be persons if X is considered severable from them; if we judge that they can, we then could call X "external." Hegel's solution is also unsatisfactory because (at least from our present vantage point) we can see that the external/internal distinction is a continuum and not a bright-line dichotomy. Both the tendency to take "external" and "internal" as premises rather than moral conclusions and the tendency to see a bright line between "external" and "internal" are traceable to the prevailing world view that posits persons as subjects in a world of objects. This world view makes it seem intuitively obvious that a thing must be either purely subjective or purely objective, and intuitively obvious into which category it falls. I am suggesting that we relinquish the subject/object dichotomy and rely instead on our best moral judgment in light of the best conception of personhood as we now understand it.[222]

B. Methods of Justifying Market-Inalienabilities

If some people wish to sell something that is identifiably personal, why not let them? In a market society, whatever some people wish to buy and others wish to sell is deemed alienable. Under these circumstances, we must formulate an affirmative case for market-inalienability, so that no one may choose to make fungible — commodify — a personal attribute, right, or thing. In this Section, I propose and evaluate three possible methods of justifying market-inalienability based on personhood: a prophylactic argument, assimilation to prohibition, and a domino theory.

The method of justification that correlates most readily with traditional liberal pluralism is a prophylactic argument. For the liberal it makes sense to countenance both selling and sharing of personal things as the holder freely chooses. If an item of property is personal, however, sometimes the circumstances under which the holder places it on the market might arouse suspicion that her act is coerced. Given that we cannot know whether anyone really intends to cut herself off from something personal by commodifying it, our suspicions might sometimes justify banning sales. The risk of harm to the seller's

---

[221] *See supra* pp. 1892–94.

[222] Of course, in suggesting that we relinquish the subject/object dichotomy, I do not mean that we should try to relinquish the common-sense understanding that there is a world of things separate from ourselves, that somehow we should to try to feel ourselves fused with our chairs or our pencils. I do mean that we need not explain our feelings of separateness from things in the world by positing the Kantian foundational apparatus of subjectivity and objectivity.

personhood in cases in which coerced transactions are permitted (especially if the thing sought to be commodified is normally very important to personhood), and the great difficulties involved in trying to scrutinize every transaction closely, may sometimes outweigh the harm that a ban would impose on would-be sellers who are in fact uncoerced. A prophylactic rule aims to ensure free choice — negative liberty — by the best possible coercion-avoidance mechanism under conditions of uncertainty.[223] This prophylactic argument is one way for a liberal to justify, for example, the ban on selling oneself into slavery. We normally view such commodification as so destructive of personhood that we would readily presume all instances of it to be coerced. We would not wish, therefore, to have a rule creating a rebuttable presumption that such transactions are uncoerced (as with ordinary contracts), nor even a rule that would scrutinize such transactions case-by-case for voluntariness, because the risk of harm to personhood in the coerced transactions we might mistakenly see as voluntary is so great that we would rather risk constraining the exercise of choice by those (if any) who really wish to enslave themselves.[224]

A liberal pluralist might use a prophylactic justification to prevent poor people from selling their children, sexual services, or body parts. The liberal would argue that an appropriate conception of coercion should, with respect to selling these things, include the desperation of poverty.[225] Poor people should not be forced to give up personal things because the relinquishment diminishes them as persons, contrary to the liberal regime of respect for persons. We should presume that such transactions are not the result of free choice.

When thus applied to coercion by poverty, the prophylactic argument is deeply troubling. If poverty can make some things nonsalable because we must prophylactically presume such sales are coerced, we would add insult to injury if we then do not provide the would-be seller with the goods she needs or the money she would have received. If we think respect for persons warrants prohibiting a mother from

---

[223] A prophylactic ban on sales would thus be a risk-of-error rule based on respect for persons. *See* Radin, *Risk-of-Error Rules and Non-Ideal Justification*, in JUSTIFICATION: NOMOS XXVIII 33 (J. Pennock & J. Chapman eds. 1986). The rules allowing all accused persons to go unpunished absent proof of guilt beyond a reasonable doubt or invalidating of all contracts involving minors as parties are examples of risk-of-error rules.

[224] In assuming that self-commodification might be acceptable but for uncertainties of knowledge and adjudication, this argument in principle admits commodification of the person. *Cf.* J. FEINBERG, *supra* note 199, at 80–81 (asserting that a blanket rule against self-enslavement might be justified by the risk of mistaken judgments of voluntariness); R. POSNER, *supra* note 8, at 238–44 (arguing that the unenforceability of a voluntary self-enslavement contract is economically explainable by the high likelihood of making a disastrous mistake).

[225] *Cf.* M. WALZER, *supra* note 42, at 102 (discussing a ban on "desperate exchanges" in the labor market).

selling something personal to obtain food for her starving children, we do not respect her personhood more by forcing her to let them starve instead. To the extent it equates poverty with coercion, the prophylactic argument requires a corollary in welfare rights. Otherwise we would be forcing the mother to endure a devastating loss in her primary relationship (with her children) rather than in the secondary one (with the personal thing) she is willing to sacrifice to protect the primary one. It is as if, when someone is coerced at gunpoint, we were to direct our moral opprobrium at the victim rather than the gun-wielder, and our enforcement efforts at preventing the victim from handing over her money rather than at preventing the gun-wielder from placing her in the situation where she must. Thus, this aspect of liberal prophylactic pluralism is hypocritical without a large-scale redistribution of wealth and power that seems highly improbable.[226] Although we may nevertheless decide to ban sales of certain personal things, the prophylactic argument, insofar as it rests on equating poverty with coercion, cannot be the reason.[227]

---

[226] If such a redistribution were to come about, poverty would no longer be presumed a coercive factor, and the prophylactic justification would be less compelling. When someone is coerced at gunpoint, the remedy is to force the gun-wielder to give back whatever was obtained under duress and to try to prevent such threats from occurring in the first place. If someone is "coerced" by poverty into selling something she would not otherwise sell, unwinding the transaction is more problematic. The buyer is not the sole cause of the seller's duress, and thus it seems unfair to take back the "goods" and let the seller keep the money. If unwinding the transaction includes restitution of the price paid, then the duress is not removed. To prevent such threats from occurring in the first place entails preventing poverty. A rule saying that those who give up anything at gunpoint will be punished would not be appropriate; thus, it seems that a rule saying that those who give up things under the "coercion" of poverty will be punished is not appropriate either. This argument can be understood as one reason why we should not necessarily consider economic need as negating free choice.

The puzzle about whether poverty can constitute coercion is a philosophical red herring that conceals a deeper problem. Insofar as preventing sales seems harmful or disempowering to poor people who otherwise would sell personal things, it is so even if we think of the choice to sell as coerced. Because allowing sales, even if we think of them as freely chosen, also seems harmful or disempowering, we are caught in a double bind, a painful dilemma of transition. *See supra* pp. 1875–77; *infra* pp. 1915–17.

[227] Although the prophylactic argument is troubling when applied to "coercion" by poverty, it retains some force with respect to coercion in general. People can be coerced by many nonmonetary factors of power others may have over them. The issue would be whether any nonmonetary factors of power that we wish to characterize as negating free choice could plausibly be presumed to result in people's attempting to sell things. The concept of coercion — in particular the issue of what factors of power we should characterize as negating free choice — is a philosophical dispute I cannot review more deeply here. *See, e.g.*, R. FLATHMAN, *supra* note 186, at 180–220; Nozick, *Coercion*, in PHILOSOPHY, SCIENCE, AND METHOD 440 (S. Morgenbesser, P. Suppes & M. White eds. 1969).

The prophylactic argument may properly recommend that trades of personal things — like the sale of family heirlooms or a homestead — be at least more closely scrutinized for voluntariness than trades of fungible things. Invalidating "contracts" produced under duress is no more than free-market hygiene. Although we do not scrutinize all contracts for duress, case-

A second method of justifying market-inalienability assimilates it to prohibition. If we accept that the commodified object is different from the "same" thing noncommodified and embedded in personal relationships, then market-inalienability is a prohibition of the commodified version, resting on some moral requirement that it not exist. What might be the basis of such a moral requirement? Something might be prohibited in its market form because it both creates and exposes wealth- and class-based contingencies for obtaining things that are critical to life itself — for example, health care — and thus undermines a commitment to the sanctity of life.[228] Another reason for prohibition might be that the use of market rhetoric, in conceiving of the "good" and understanding the interactions of people respecting it, creates and fosters an inferior conception of human flourishing. For example, we accept an inferior conception of personhood (one allied to the extreme view of negative freedom) if we suppose people may freely choose to commodify themselves.[229]

The prohibition argument — that commodification of things is bad in itself, or because these things are not the "same" things that would be available to people in nonmarket relationships — leads to universal noncommodification. If commodification is bad in itself it is bad for everything. Any social good is arguably "different" if not embedded in a market society.[230] To restrict the argument in order to permit pluralism, we have to accept either that certain things are the "same" whether or not they are bought and sold, and others are "different," or that prohibiting the commodified version morally matters only for certain things, but not for all of them. At present we tend to think that nuts and bolts are pretty much the "same" whether commodified or not, whereas love, friendship, and sexuality are very "different"; we also tend to think that trying to keep society free of commodified love, friendship, and sexuality morally matters more than trying to keep it free of commodified nuts and bolts.[231]

A third method of justifying market-inalienability, the domino theory, envisions a slippery slope leading to market domination. The

---

by-case analysis of trades of things that are usually personal could be mandated by the conviction that respect for personhood requires individualized attention.

[228] *See* M. Shapiro, Regulation as Language: Communicating Values by Reducing the Contingencies of Choice 14–15, 25–28, 53–57 (unpublished manuscript on file at Harvard Law School Library).

[229] *See supra* pp. 1898–903, 1905.

[230] This point is made by writers as disparate as Georg Lukács, cited in note 88 above, and Peter Singer, cited in note 29 above.

[231] To this the universal noncommodifier would no doubt respond that commodified nuts and bolts are produced by commodified labor, and that prohibiting commodified labor morally matters as much as prohibiting commodified love, friendship, and sexuality. She might further respond that commodification of their labor forces workers to experience only the commodified versions of love, friendship, and sexuality.

domino theory assumes that for some things, the noncommodified version is morally preferable; it also assumes that the commodified and noncommodified versions of some interactions cannot coexist. To commodify some things is simply to preclude their noncommodified analogues from existing. Under this theory, the existence of some commodified sexual interactions will contaminate or infiltrate everyone's sexuality so that all sexual relationships will become commodified. If it is morally required that noncommodified sex be possible, market-inalienability of sexuality would be justified. This result can be conceived of as the opposite of a prohibition: there is assumed to exist some moral requirement that a certain "good" be socially available. The domino theory thus supplies an answer (as the prohibition theory does not) to the liberal question why people should not be permitted to choose both market and nonmarket interactions: the noncommodified version is morally preferable when we cannot have both.

We can now see how the prohibition and domino theories are connected. The prohibition theory focuses on the importance of excluding from social life commodified versions of certain "goods" — such as love, friendship, and sexuality — whereas the domino theory focuses on the importance for social life of maintaining the noncommodified versions. The prohibition theory stresses the wrongness of commodification — its alienation and degradation of the person — and the domino theory stresses the rightness of noncommodification in creating the social context for the proper expression and fostering of personhood. If one explicitly adopts both prongs of this commitment to personhood, the prohibition and domino theories merge.[232]

The argument that market-inalienabilities are necessary to encourage altruism relies upon the domino theory. With regard to human blood, Richard Titmuss argues that a regime permitting only donation fosters altruism.[233] The altruistic experience of the donor in being responsible (perhaps) for saving a stranger's life is said to bring us closer together, cementing our community in a way that buying and selling cannot.[234] The possibility of reciprocity is also a part of this cementing process, because a donor's sense of obligation could be partially founded on the recognition that she could be a recipient some day. From the recipient's perspective, it is said that knowing one is

---

[232] In fact, utopian noncommodifiers, who think that commodification is inherently wrong, also tend to think that commodified and noncommodified forms of human interactions cannot coexist. In his view that "bourgeois property" cannot coexist with other kinds of property, Marx may be understood to have meant that market and nonmarket forms cannot coexist. *See supra* note 144; *see also* G. LUKÁCS, *supra* note 88.

[233] *See* R. TITMUSS, *supra* note 6. Peter Singer uses the form of argument I call the domino theory in his defense of Titmuss against the liberal view that both gifts and sales should be permitted. *See generally* Singer, *supra* note 29.

[234] *See* R. TITMUSS, *supra* note 6, at 237–46; Singer, *supra* note 29.

dependent on others' altruism rather than on one's own wealth creates solidarity and interdependence, and that this knowledge of dependence better preserves and expresses the ideal of sanctity of life. But why do we need to forbid sales to preserve opportunities for altruism for those who wish to give? In a gifts-only regime, a donor's gift remains nonmonetized, whereas if both gifts and sales are permitted, the gift has a market value. This market value undermines our altruism and discourages us from giving, the argument runs, because our gift is now equivalent merely to giving fifty dollars (or whatever is the market price of a pint of blood) to a stranger, rather than life or health.

The "domino" part of this argument — that once something is commodified for some it is willy-nilly commodified for everyone — posits that once market value enters our discourse, market rhetoric will take over and characterize every interaction in terms of market value. If this is true, some special things (for example, blood) must be completely noncommodified if altruism is to be possible.[235] But the feared domino effect of market rhetoric need not be true. To suppose that it must necessarily be true seems to concede to universal commodification the assumption that thinking in money terms comes "naturally" to us.[236] Most people would probably think the assumption false in light of their common experience. For example, many people value their homes or their work in a nonmonetary way, even though those things also have market value.[237]

Rather than merely assuming that money is at the core of every transaction in "goods," thereby making commodification inevitable and phasing out the non-commodified version of the "same" thing (or the nonmarket aspects of sale transactions), we should evaluate the domino theory on a case-by-case basis. We should assess how important it is to us that any particular contested thing remain available in a noncommodified form and try to estimate how likely it is that allowing market transactions for those things would engender a domino effect and make the nonmarket version impossible. This might involve judging how close to universal commodification our consciousness really is, and how this consciousness would affect the particular thing in question.

---

[235] In my view, the argument against commodification of blood on the ground that it would foreclose a necessary opportunity for altruism does not succeed, because the argument is too general to single out blood or any small group of particular things. *See* Radin, *supra* note 42.

[236] The assumption is a concession to universal commodification if it means that thinking in money terms comes naturally to people *sub specie aeternitatis*. But noncommodifiers might assume that thinking in money terms comes naturally to people who live in a commodified social order. This assumption expresses the link between rhetoric and the world, discussed in Section B of Part III above. My argument is that it should be evaluated more particularly, not that it should be ignored.

[237] *See infra* pp. 1918–21.

C. The Problem of Nonideal Evaluation

One ideal world would countenance no commodification; another would insist that all harms to personhood are unjust; still another would permit no relationships of oppression or disempowerment. But we are situated in a nonideal world of ignorance, greed, and violence; of poverty, racism, and sexism. In spite of our ideals, justice under nonideal circumstances, pragmatic justice, consists in choosing the best alternative now available to us. In doing so we may have to tolerate some things that would count as harms in our ideal world. Whatever harms to our ideals we decide we must now tolerate in the name of justice may push our ideals that much farther away.[238] How are we to decide, now, what is the best transition toward our ideals, knowing that our choices now will help to reconstitute those ideals?

The possible avenues for justifying market-inalienability must be reevaluated in light of our nonideal world. In light of the desperation of poverty, a prophylactic market-inalienability may amount merely to an added burden on would-be sellers; under some circumstances we may judge it, nevertheless, to be our best available alternative. We might think that both nonmarket and market interactions can exist in some situations without a domino effect leading to a more commodified order, or we might think it is appropriate to risk a domino effect in light of the harm that otherwise would result to would-be sellers. We might find prohibition of sales not morally warranted, on balance, in some situations, unless there is a serious risk of a domino effect. These will be pragmatic judgments.

1. The Double Bind. — Often commodification is put forward as a solution to powerlessness or oppression, as in the suggestion that

---

[238] Thus, the problem of justice under nonideal circumstances is connected to the transition dilemma for social progress that I discussed earlier. *See supra* pp. 1875–77. Although I shall not try to do so in this Article, we are now in a better position to evaluate a narrower transition problem: whether changes, though necessary for progress, require compensation of individuals who consider themselves harmed by them. *Cf.* Kaplow, *An Economic Analysis of Legal Transitions*, 99 HARV. L. REV. 509 (1986) (arguing that compensation is generally not required). For example, are holders of fungible property wronged (and thereby perhaps entitled to compensation) when it is partially decommodified? If the question is analyzed in moral terms other than those of the market, it makes a difference what kind of thing we are talking about. If the attribute, right, or thing is fungible to those who consider themselves losers, its loss is appropriately considered monetary; if it is not fungible to those who benefit from the change, our best theory of personhood might tell us that it ought not to be commodified. In this case decommodification corrects a wrong to personhood at the expense only of fungible interests held wrongfully. Thus, when some things emerge as more important than ordinary fungible property to people and to society as a whole, the diminution of their fungible value to would-be sellers might be seen as an acceptable responsibility of citizenship. It appears at least that to pay compensation — for example, to those who lost money when slaves were emancipated or when child labor was prohibited — would deny that the thing had been decommodified, treating it as fungible even while declaring that it is not.

women be permitted to sell sexual and reproductive services.[239] But is women's personhood injured by allowing or by disallowing commodification of sex and reproduction? The argument that commodification empowers women is that recognition of these alienable entitlements will enable a needy group — poor women — to improve their relatively powerless, oppressed condition, an improvement that would be beneficial to personhood. If the law denies women the opportunity to be comfortable sex workers and baby producers instead of subsistence domestics, assemblers, clerks, and waitresses — or pariahs (welfare recipients) and criminals (prostitutes) — it keeps them out of the economic mainstream and hence the mainstream of American life.

The rejoinder is that, on the contrary, commodification will harm personhood by powerfully symbolizing, legitimating, and enforcing class division and gender oppression. It will create the two forms of alienation that correlate with commodification of personal things.[240] Women will partly internalize the notion that their persons and their attributes are separate, thus creating the pain of a divided self. To the extent that this self-conception is not internalized, women will be alienated from the dominant order that, by allowing commodification, sees them in this light. Moreover, commodification will exacerbate, not ameliorate, oppression and powerlessness, because of the social disapproval connected with marketing one's body.[241]

But the surrejoinder is that noncommodification of women's capabilities under current circumstances represents not a brave new world of human flourishing, but rather a perpetuation of the old order that submerges women in oppressive status relationships, in which personal identity as market-traders is the prerogative of males. We cannot make progress toward the noncommodification that might exist under ideal conditions of equality and freedom by trying to maintain noncommodification now under historically determined conditions of inequality and bondage.

These conflicting arguments illuminate the problem with the prophylactic argument for market-inalienability.[242] If we now permit commodification, we may exacerbate the oppression of women — the suppliers. If we now disallow commodification — without what I have called the welfare-rights corollary, or large-scale redistribution

---

[239] Although in the text I pursue its application to poor women, it should be evident that the double bind is broader in scope. For example, it also complicates the problem of whether people should be allowed to sell their organs. *See supra* note 23.

[240] *See supra* p. 1907.

[241] If marketing one's body is an available option, then those who fail to commodify themselves to feed their families might be thought blameworthy as well. *See* M. Shapiro, *supra* note 228, at 55.

[242] *See supra* pp. 1910–11.

of social wealth and power[243] — we force women to remain in circumstances that they themselves believe are worse than becoming sexual commodity-suppliers. Thus, the alternatives seem subsumed by a need for social progress, yet we must choose some regime now in order to make progress. This dilemma of transition is the double bind.[244]

The double bind has two main consequences. First, if we cannot respect personhood either by permitting sales or by banning sales, justice requires that we consider changing the circumstances that create the dilemma. We must consider wealth and power redistribution. Second, we still must choose a regime for the meantime, the transition, in nonideal circumstances. To resolve the double bind, we have to investigate particular problems separately; decisions must be made (and remade) for each thing that some people desire to sell.

If we have reason to believe with respect to a particular thing that the domino theory might hold — commodification for some means commodification for all — we would have reason to choose market-inalienability. But the double bind means that if we choose market-inalienability, we might deprive a class of poor and oppressed people of the opportunity to have more money with which to buy adequate food, shelter, and health care in the market, and hence deprive them of a better chance to lead a humane life. Those who gain from the market-inalienability, on the other hand, might be primarily people whose wealth and power make them comfortable enough to be concerned about the inroads on the general quality of life that commodification would make. Yet, taking a slightly longer view, commodification threatens the personhood of everyone, not just those who can now afford to concern themselves about it. Whether this elitism in market-inalienability should make us risk the dangers of commodification will depend upon the dangers of each case.

2. Incomplete Commodification. — One way to mediate the dilemma is through what I shall call incomplete commodification. Un-

---

[243] *See id.*

[244] In the struggle for social justice for women, the double bind is pervasive. Is marriage to be considered a contract in which certain distributions of goods are agreed to between autonomous bargaining agents? Upon divorce, such a conception of marriage makes it difficult for oppressed women who have not bargained effectively to obtain much. Or is marriage to be considered a noncontractual sharing status in which the partners' contributions are not to be monetized? Upon divorce, such a conception makes it difficult for oppressed women who have contributed unmonetized services to their husbands' advantage to obtain much. Unmonetized sharing is hypocritical under present social circumstances, say feminist theorists, yet in a better world many feminists would prefer unmonetized sharing to commodification. The idea of contractual autonomy is more attractive than being submerged in a status that gives all power to men, yet the autonomy is often illusory, and the reinforcement of individualist bargaining models of social interaction is contrary to feminist ideals for a better world. *See* sources cited *infra* notes 259-60.

der nonideal circumstances the question whether market-inalienability can be justified is more complicated than a binary decision between complete commodification and complete noncommodification.[245] Rather, we should understand there to be a continuum reflecting degrees of commodification that will be appropriate in a given context. An incomplete commodification — a partial market-inalienability[246] — can sometimes substitute for a complete noncommodification that might accord with our ideals but cause too much harm in our nonideal world.

Before considering examples, it may be helpful to distinguish two aspects of incomplete commodification: participant and social.[247] The participant aspect draws attention to the meaning of an interaction for those who engage in it. For many interactions in which money changes hands, market rhetoric cannot capture this significance. In other words, market and non market aspects of an interaction coexist: although money changes hands, the interaction also has important nonmonetizable personal and social significance. The social aspect of incomplete commodification draws attention instead to the way society as a whole recognizes that things have nonmonetizable participant significance by regulating (curtailing) the free market.

Work and housing are possible examples of incomplete commodification. With respect to the participant aspect, consider that for many of us, work is not only the way we make our living, but also a part of ourselves. What we hope to derive from our work, and the personal importance we attach to it, are not understandable entirely in money terms, even though we demand and accept money. These ideals about work seem to be part of our conception of human flourishing, and thus the loss of this personal aspect of work would be considered inhumane.[248] Consider also our attachment of meaning to housing. Although a house has market value and we can express our investment

---

[245] To think of the problem as simply drawing a boundary line between a completely nonmarket realm and a laissez-faire market realm is under present circumstances to understand (with liberal pluralism) the market as the presumptive norm. *See supra* note 153.

[246] The conventional term for incomplete commodification (partial market-inalienability) is, of course, restraint upon alienation.

[247] These aspects correspond to the two ways in which universal commodification views the human world as completely commodified. In the participant aspect — the meaning of interactions to the individuals involved in them — all human interactions are viewed as trades resulting in monetizable gains and losses. In the social aspect — the community response to this meaning — all things are presumptively to be traded in a free market.

[248] Thus, to think of our labor power only as a commodity separate from ourselves, is, as Marx thought, to do violence to our ideal of personhood. In supposing that for some of us work is incompletely commodified, I am supposing — perhaps contrary to Marxists — that unalienated work exists to some extent. *See supra* pp. 1871–72. I am not supposing that no alienated labor exists, nor am I supposing that unalienated work is not correlated with class. For further discussion of incomplete commodification of work, see Radin, cited in note 42 above.

in terms of dollars, there is a nonmonetizable, personal aspect to many people's relationships with their homes.

With respect to the social aspect of incomplete commodification, consider the regulation of labor. Although work has not been fully decommodified, it is incompletely commodified through collective bargaining, minimum wage requirements, maximum hour limitations, health and safety requirements, unemployment insurance, retirement benefits, prohibition of child labor, and antidiscrimination requirements.[249] Consider also the regulation of residential tenancies. Rent control, habitability requirements, restrictions upon termination of tenancies, and antidiscrimination requirements can all be seen as indicia of incomplete commodification.

When we see these regulations as reflecting incomplete commodification, we progress toward conceiving of work and housing in other than market rhetoric.[250] In this view, work and housing are not conceived of as completely monetizable and fungible objects of exchange that are separated from persons, because to conceive of them in such a way is to adopt an inferior conception of human flourishing.[251] These forms of regulation should instead be seen as an effort

---

[249] Karl Polanyi noted the partial decommodification of labor:

> To argue that social legislation, factory laws, unemployment insurance, and, above all, trade unions have not interfered with the mobility of labor and the flexibility of wages, as is sometimes done, is to imply that those institutions have entirely failed in their purpose, which was exactly that of interfering with the laws of supply and demand in respect to human labor, and removing it from the orbit of the market.

K. POLANYI, *supra* note 139, at 177.

[250] In examining the connection between personhood and one's work or home, the discussion in the text links incomplete commodification of labor and housing to fostering freedom, identity, and especially contextuality. Perhaps labor regulation could also be seen as designed to check a domino effect: complete commodification of people's labor may (as Marx thought) lead to commodification of workers. *See supra* pp. 1871–73.

There are other, more market-oriented ways of looking at regulation. From the point of view of universal commodification, these regulations would be unjustified unless they promote efficiency, and they have not been readily seen as efficiency-enhancing. *See, e.g.*, R. POSNER, *supra* note 8, §§ 11.6–11.7, at 308–15 (discussing inefficiencies resulting from regulation of the employment relationship); *id.* § 16.6, at 445–48 (discussing inefficiencies resulting from housing code enforcement); Epstein, *A Common Law for Labor Relations: A Critique of the New Deal Labor Legislation*, 92 YALE L.J. 1357 (1983); Epstein, *In Defense of the Contract at Will*, 51 U. CHI. L. REV. 947 (1984); *Hirsch*, supra note 107. *But cf.* Donohue, *Is Title VII Efficient?*, 134 U. PA. L. REV. 1411 (1986) (using an economic model to argue that antidiscrimination legislation might enhance economic efficiency). These regulations may also be seen as examples of wealth redistribution under liberal welfare rights, but this understanding also tends toward commodification: it assimilates work and personal property to fungible wealth of the holders, merely asserting that the holders are entitled to a minimum amount of such wealth even if that requires redistribution. For a discussion of welfare rights and personhood in the context of housing, see Radin, *Residential Rent Control*, cited in note 220 above.

[251] Insofar as we do see regulation as the social aspect of incomplete commodification, moral reasoning and not market failure will be the focus of debates over its proper extent. We would justify regulation of interactions involving buying and selling these things by referring to their

to take into account workers' and tenants' personhood, to recognize and foster the nonmarket significance of their work and housing.

Regulation of residential tenancies can be seen as connected to identity and contextuality: attempting to make possible and protect the constituting of one's personhood in one's home, and one's continuity of residence there, because the home is a justifiable kind of personal property.[252] Regulation can be seen as attempting to ensure that tenants are not forced to move from their homes for ideological, discriminatory, or arbitrary reasons, or by a sudden rise in market prices, and to ensure that rental housing is decent to live in and a decent place for family life.

Regulation of work can be seen as attempting to make more possible the realization of personal ideals about work, which are related to human flourishing: a self-conception inseparable from one's work (contextuality), continuity of work (identity), and control over one's own work (freedom). Regulation can be seen as attempting to ensure that employees are not forced to leave their jobs for ideological, discriminatory, or arbitrary reasons; to ensure that the workplace is safe, and free from sexual or racial harassment; and to ensure that employees have some say in workplace decisions, and the opportunity to understand how their work is helpful or significant to other people.[253] Although complete decommodification of work or housing is

---

closeness to personhood, not to inefficiencies caused by transaction costs. If we think that because of their desperate poverty and the pricing policies of landlords and employers, tenants and workers would not wish to have the regulations, that is, would choose complete commodification, then that places us in the double bind. *See supra* pp. 1915–17. The regulations are not thereby rendered unjustified, however, if they are now our best available alternative.

[252] Thus, the category of personal property may be seen as related to incomplete commodification. *See supra* notes 115–17. For those things that we accept as being appropriately identified with the person, a range of protections exists to shield them from market forces and wrongful treatment as fungible. The ability to establish oneself in relationship with things is promoted by the social aspect of incomplete commodification; once the relationship is established, the thing is personal.

[253] Other regulation of labor and rental housing can also be seen as partial decommodification based on personhood. For example, there is some tendency toward recognizing job-tenure rights, *see, e.g.*, *Individual Rights in the Workplace: The Employment-At-Will Issue*, 16 U. MICH. J.L. REF. 199 (1983); Hermann & Sor, *Property Rights in One's Job: The Case for Limiting Employment-At-Will*, 24 ARIZ. L. REV. 763 (1982), and a parallel tendency toward recognizing tenants' tenure rights, *see, e.g.*, RESTATEMENT (SECOND) OF PROPERTY § 14.8 (1977) (prohibiting retaliatory evictions in residential housing); Uniform Residential Landlord and Tenant Act § 5.101 (1972) (same); N.J. STAT. ANN. § 2A:18–61.1 (West Supp. 1976) (limiting the termination of residential tenancies to enumerated grounds of "good cause"); Baar, *Guidelines for Drafting Rent Control Laws: Lessons of a Decade*, 35 RUTGERS L. REV. 723, 833–35 (1983) (noting that eviction control accompanies rent control). Partial decommodification of labor is proceeding further with the advent of comparable worth (a form of just price regulation), *see, e.g.*, COMPARABLE WORTH AND WAGE DISCRIMINATION: TECHNICAL POSSIBILITIES AND POLITICAL REALITIES (H. Remick ed. 1984); Weiler, *The Wages of Sex: The Uses and Limits of Comparable Worth*, 99 HARV. L. REV. 1728 (1986), and partial decommodification of rental housing is proceeding further with rent control (also a form of just price regulation). *See* Baar, *supra*, at 725–26 & n.1 (estimating that approximately 10% of all privately owned residential units in the

not now possible, these incomplete commodifications can be seen as responses in our nonideal world to the harm to personhood caused by complete commodification of work and housing.

D. Evolutionary Pluralism Applied: Problems of Sexuality and Reproductive Capacity

I now offer thoughts on how the analysis that I recommend might be brought to bear on a set of controversial market-inalienabilities. It is not my purpose to try to provide the detailed, practical evaluation that is needed, but only to sketch its general contours. The example I shall pursue is the contested commodification of aspects of sexuality and reproductive capacity: the issues of prostitution, baby-selling, and surrogacy.[254] I conclude that market-inalienability is justified for baby-selling and also — provisionally — for surrogacy, but that prostitution should be governed by a regime of incomplete commodification.

Assuming that our ideal of personhood includes the ideal of sexual interaction as equal nonmonetized sharing, we might imagine that the "good" commodified sexuality ought not to exist: that sexual activity should be market-inalienable.[255] But perhaps prohibition of the sale of sexual services, if it aims to preserve sexuality as nonmonetized sharing, is not justified under current circumstances, because sex is already commodified.[256] Moreover, in our nonideal world, market-inalienability — especially if enforced through criminalization of sales — may cause harm to ideals of personhood instead of maintaining

---

United States are subject to some form of rent control); Radin, *Residential Rent Control*, *supra* note 220.

[254] These are some of the central cases of contested commodification. For other cases to which market-inalienability might be extended, see notes 27–33, 235 above. There are occasionally market-inalienabilities attached to things that seem "external," in the sense of not being closely related to personhood as now conceived. Many of these invite us to consider human personhood in a broad ecological context. Examples are legislation banning hunting and trade in artifacts of endangered species. *See, e.g.*, Andrus v. Allard, 444 U.S. 51, 64–68 (1979) (upholding a market inalienability in the Eagle Protection Act, 16 U.S.C. § 668(a) (1982), against challenge as a taking); *cf.* Rose-Ackerman, *supra* note 10, at 943 (discussing these inalienabilities from an economic efficiency perspective).

[255] I am confining the present discussion to traditional male-female prostitution because I am considering a set of would-be commodities that women would control. A different nonideal moral analysis will no doubt be required for other forms.

[256] Legalized prostitution has existed in many places, and there has always been a large black market of which everyone is well aware. That those who purchase prostitutes' services are often not prosecuted seems to indicate that commodification of sexuality, at least by the purchasers, is tolerated. For various views on commodification and prostitution, see, for example, Jaggar, *Prostitution*, in THE PHILOSOPHY OF SEX (A. Soble ed. 1980); Richards, *Commercial Sex and the Rights of the Person: A Moral Argument for the Decriminalization of Prostitution*, 127 U. PA. L. REV. 1195 (1979); and sources cited in note 31 above.

and fostering them, primarily because it exacerbates the double bind.[257] Poor women who believe that they must sell their sexual services to survive are subject to moral opprobrium, disease, arrest, and violence. The ideal of sexual sharing is related to identity and contextuality, but the identity of those who sell is undermined by criminalization and powerlessness, and their ability to develop and maintain relationships is hurt by these circumstances.

Nevertheless, despite the double bind and the harms of the black market to prostitutes, fear of a domino effect could perhaps warrant market-inalienability as an effort to ward off conceiving of all sexuality as fungible. Many people would say, however, that the known availability of commodified sex by itself does not render nonfungible sexual interactions impossible or even more difficult, and that the prevalence of ideals of interpersonal sexual sharing in spite of the widespread association of sex and money is proof that the domino effect in rhetoric is not to be feared. But we must evaluate the seriousness of the risk if commodification proceeds. What if sex were fully and openly commodified? Suppose newspapers, radio, TV, and billboards advertised sexual services as imaginatively and vividly as they advertise computer services, health clubs, or soft drinks. Suppose the sexual partner of your choice could be ordered through a catalog, or through a large brokerage firm that has an "800" number, or at a trade show, or in a local showroom. Suppose the business of recruiting suppliers of sexual services was carried on in the same way as corporate headhunting or training of word-processing operators. A change would occur in everyone's discourse about sex, and in particular about women's sexuality. New terms would emerge for particular gradations of market value, and new discussions would be heard of particular abilities or qualities in terms of their market value. With this change in discourse would come a change in everyone's experience.[258] The open market might render subconscious valuation of women (and perhaps everyone) in sexual dollar value impossible to avoid. It might make the ideal of nonmonetized sharing impossible. Thus, the argument for noncommodification of sexuality based on the domino effect, in its strongest form, is that we do not wish to unleash market forces onto the shaping of our discourse regarding sexuality and hence onto our very conception of sexuality and our sexual feelings.

This domino argument assumes that nonmonetized equal sharing relationships are the norm or are at least attainable. That assumption is now contested. Some feminists argue that male-female sexual relationships that actually instantiate the ideal of equal sharing are under current social circumstances rare or even impossible.[259] Ac-

---

[257] *See supra* pp. 1915–17.

[258] This is the lesson of the effect of rhetoric on our world. *See supra* pp. 1881–85.

[259] *See, e.g.*, C. MACKINNON, FEMINISM UNMODIFIED: DISCOURSES ON LIFE AND LAW

cording to this view, moreover, women are oppressed by this ideal because they try to understand their relationships with men in light of it, and conceal from themselves the truth about their own condition. They try to understand what they are doing as giving, as equal sharing, while their sexuality is actually being taken from them. If we believe that women are deceived (and deceiving themselves) in this way, attempted noncommodification in the name of the ideal may be futile or even counterproductive. Noncommodification under current circumstances is part of the social structure that perpetuates false consciousness about the current role of the ideal. Some feminists also argue that many male-female sexual relationships are (unequal) economic bargains, not a context in which equal sharing occurs.[260] If that is true, attempted noncommodification means that prostitutes are being singled out for punishment for something pervasive in women's condition, and that they are being singled out because their class or race forecloses more socially accepted forms of sexual bargaining. This returns us to the double bind.

Perhaps the best way to characterize the present situation is to say that women's sexuality is incompletely commodified. Many sexual relationships may have both market and nonmarket aspects: relationships may be entered into and sustained partly for economic reasons and partly for the interpersonal sharing that is part of our ideal of human flourishing. Even if under current circumstances the ideal misleads us into thinking that unequal relationships are really equal, it seems that the way out of such ideological bondage is not to abandon the ideal, but rather to pursue it in ways that are not harmful under

---

(1987); Gottlieb, *The Political Economy of Sexuality*, 16 REV. RADICAL POL. ECON. 143 (1984); Hantzis, *Is Gender Justice a Completed Agenda?* (Book Review), 100 HARV. L. REV. 690 (1987); MacKinnon, *Feminism, Marxism, Method and the State: Toward Feminist Jurisprudence*, 8 SIGNS: J. WOMEN CULTURE & SOC. 635 (1983); MacKinnon, *Feminism, Marxism, Method, and the State: An Agenda for Theory*, 7 SIGNS: J. WOMEN CULTURE & SOC. 515 (1982).

[260] *See, e.g.*, A. JAGGAR, FEMINIST POLITICS AND HUMAN NATURE (1983); P. ROOS, GENDER AND WORK 119–54 (1985); Rubin, *The Traffic in Women*, in TOWARD AN ANTHROPOLOGY OF WOMEN 157 (R. Reiter ed. 1975). Insistence on continued noncommodification of homemaker services of a wife is also problematic. The context of current sexual politics makes both commodification and noncommodification seem generally disempowering to women. Assimilation to the market paradigm seems defeating for personhood, relationships, and political identity, but given economic and cultural realities, so does continued insistence on a realm of nonmarket interpersonal sharing.

The additional argument that the commodity form of a thing might drive out the noncommodified version of the "same" thing does not seem at present a great threat to nonmarketized homemaker services. A domestic services market (though not one that is in full bloom) does coexist with a parallel class of unpaid providers. It does not appear that, as a result, we have implicitly come to think of homemaker services in market rhetoric. And if we had — here is the double bind again — many women would be better off at divorce, when money is all that is left at stake. *Cf.* L. WEITZMAN, THE DIVORCE REVOLUTION 323–401 (1985) (describing the disastrous economic consequences to women and children of the present system of divorce).

these nonideal circumstances. Market-inalienability seems harmful, not only because it might be ideologically two-edged, but also because of the double bind.. Yet complete commodification, if any credence is given to the feared domino effect, may relinquish our conception of sexuality entirely.

The issue thus becomes how to structure an incomplete commodification that takes account of our nonideal world, yet does not foreclose progress to a better world of more equal power (and less susceptibility to the domino effect of market rhetoric). I think we should now decriminalize the sale of sexual services in order to protect poor women from the degradation and danger either of the black market or of other occupations that seem to them less desirable. At the same time, in order to check the domino effect, we should prohibit the capitalist entrepreneurship that would operate to create an organized market in sexual services even though this step would pose enforcement difficulties. It would include, for example, banning brokerage (pimping) and recruitment.[261] It might also include banning

---

[261] In another double bind situation, the sale of human organs, California has imposed an analogous incomplete commodification, providing that patients may sell their organs but criminalizing brokering. *See* CAL. PENAL CODE § 367f(e) (West 1986). In structuring an incomplete commodification for sexual activity, another important issue is whether contracts to sell sexual services should be enforced. The usual reason given for precluding specific performance of personal service agreements is that to force someone to perform smacks of slavery. If sexual service contracts were to be specifically performed, persons would be forced to yield their bodily integrity and freedom; this is commodification of the person. Suppose, then, that we decide to preclude specific performance but allow a damage remedy. Enforceable contracts might make the "goods" command higher prices. This is on the procommodification side of the double bind. The other side is that having to pay damages for deciding not to engage in sex with someone seems very harmful to the ideal of sexuality as integral to personhood, and it seems that determining the amount of damages due is tantamount to complete commodification. Granting a damage remedy requires an official entity to place a fungible value on the "goods"; commodification is thus officially imposed. Thus, we should make such contracts unenforceable, denying the most important factor of commodification — enforceable free contract. We could either provide for restitution if the woman reneges or let losses lie. If we let losses lie, we preclude any increased domino effect that official governmental (court) pronouncements about commodified sexuality might cause. But letting losses lie would also allow men to take and not pay when women are ignorant or powerless enough to fail to collect in advance.

Similar two-edged results are reached by the doctrine of nonenforcement of illegal contracts, under which contracts to render sexual services are currently unenforceable because of the illegality of prostitution. *See, e.g.*, CORBIN ON CONTRACTS § 1476 (1962) (stating that a contract in furtherance of immorality is void). Under this doctrine there can be a no-win situation for women in situations of nonmarital cohabitation, if the relationship is either construed as too close to countenance commodification, thus becoming an illegal contract for sexual services, or too distant to infer it, thus becoming one in which no contract in fact was made. *See* Dalton, *An Essay in the Deconstruction of Contract Doctrine*, 94 YALE L.J. 997, 1101–02 (1985). This seems to be an instance of the double bind: women who lack power are oppressed both by the noncommodification interpretation and by the commodification interpretation.

advertising. Trying to keep commodification of sexuality out of our discourse by banning advertising does have the double bind effect of failing to legitimate the sales we allow, and hence it may fail to alleviate significantly the social disapproval suffered by those who sell sexual services. It also adds "information costs" to their "product," and thus fails to yield them as great a "return" as would the full-blown market. But these nonideal effects must be borne if we really accept that extensive permeation of our discourse by commodification-talk would alter sexuality in a way that we are unwilling to countenance.[262]

A different analysis is warranted for baby-selling. Like relationships of sexual sharing, parent-child relationships are closely connected with personhood, particularly with personal identity and contextuality. Moreover, poor women caught in the double bind raise the issue of freedom: they may wish to sell a baby on the black market,[263] as they may wish to sell sexual services, perhaps to try to provide adequately for other children or family members.[264] But the double bind is not the only problem of freedom implicated in baby-selling. Under a market regime, prostitutes may be choosing to sell their sexuality, but babies are not choosing for themselves that under current nonideal circumstances they are better off as commodities. If we permit babies to be sold, we commodify not only the mother's (and father's) baby-making capacities — which might be analogous to commodifying sexuality — but we also conceive of the baby itself in market rhetoric. When the baby becomes a commodity, all of its personal attributes — sex, eye color, predicted I.Q., predicted height, and the like — become commodified as well.[265] This is to conceive of potentially all personal attributes in market rhetoric, not merely those of sexuality.

---

[262] There may be cases in which it is clearer than it is with regard to sexuality that market-inalienability is presently unjustified. A possible example is "amateur" athletics, in particular the services of collegiate football players. *See supra* note 32. At present, the issue seems to be not whether the accoutrements of capitalism can be kept out of athletics, but instead whether the suppliers of services are to be allowed a share of the take. This situation is analogous to prostitution in that some of the suppliers may be members of oppressed classes, but disanalogous in that commodifying these services is not at present connected with any particular moral opprobrium. Moreover, although the noncommodified form of athletic endeavor may be part of an interaction that is important for certain kinds of participatory bonding, that interaction is not at present as central to personhood as sexual interaction.

[263] *See generally* N. BAKER, BABYSELLING: THE SCANDAL OF BLACK-MARKET ADOPTIONS (1978).

[264] *See* N. BAKER, *supra* note 263, at 43 (suggesting that most natural mothers who give up babies for adoption on the black market are 13- to 14-year-old girls).

[265] Hence, as Posner says, there would be "good" babies (and presumably "bad" babies). *See* Posner, *supra* note 8, at 142 (quoted in note 51 above). As a result, boy babies might be "worth" more than girl babies; white babies might be "worth" more than nonwhite babies.

Moreover, to conceive of infants in market rhetoric is likewise to conceive of the people they will become in market rhetoric, and to create in those people a commodified self-conception.

Hence, the domino theory has a deep intuitive appeal when we think about the sale of babies. An idealist might suggest, however, that the fact that we do not now value babies in money suggests that we would not do so even if babies were sold. Perhaps babies could be incompletely commodified, valued by the participants to the interaction in a nonmarket way, even though money changed hands. Although this is theoretically possible, it seems too risky in our nonideal world.[266] If a capitalist baby industry were to come into being, with all of its accompanying paraphernalia, how could any of us, even those who did not produce infants for sale, avoid subconsciously measuring the dollar value of our children? How could our children avoid being preoccupied with measuring their own dollar value? This makes our discourse about ourselves (when we are children) and about our children (when we are parents) like our discourse about cars. Seeing commodification of babies as an inevitable and grave injury to personhood appears rather easy.[267] In the worst case, market rhetoric could create a commodified self-conception in everyone, as the result of commodifying every attribute that differentiates us and that other people value in us, and could destroy personhood as we know it.

I suspect that an intuitive grasp of the injury to personhood involved in commodification of human beings is the reason many people lump baby-selling together with slavery.[268] But this intuition can be misleading. Selling a baby, whose personal development requires caretaking, to people who want to act as the caretakers is not the

---

[266] Perhaps we should separately evaluate the risk in the cases of selling "unwanted" babies and selling babies commissioned for adoption or otherwise "produced" for sale. The risk of complete commodification may be greater if we officially sanction bringing babies into the world for purposes of sale than if we sanction accepting money once they are already born. It seems such a distinction would be quite difficult to enforce, however, because nothing prevents the would-be seller from declaring any child to be "unwanted." Thus, permitting the sale of any babies is perhaps tantamount to permitting the production of them for sale.

[267] As Lewis Hyde recounts:

> In 1980 a New Jersey couple tried to exchange their baby for a secondhand Corvette worth $8,800. The used-car dealer (who had been tempted into the deal after the loss of his own family in a fire) later told the newspapers why he changed his mind: "My first impression was to swap the car for the kid. I knew moments later that it would be wrong — not so much wrong for me or the expense of it, but what would this baby do when he's not a baby anymore? How could this boy cope with life knowing he was traded for a car?"

L. HYDE, *supra* note 101, at 96 n.* (1979).

[268] It is sometimes said that baby-selling violates the thirteenth amendment. *See, e.g.*, Holder, *Surrogate Motherhood: Babies for Fun and Profit*, 12 LAW, MED. & HEALTH CARE 115 (1984). For a summary of various arguments leveled against baby-selling, see Prichard, cited in note 9 above.

same thing as selling a baby or an adult to people who want to act only as users of her capacities. Moreover, if the reason for our aversion to baby-selling is that we believe it is like slavery, then it is unclear why we do not prohibit baby-giving (release of a child for adoption) on the ground that enslavement is not permitted even without consideration. We might say that respect for persons prohibits slavery but may require adoption in cases in which only adoptive parents will treat the child as a person, or in the manner appropriate to becoming a person. But this answer is still somewhat unsatisfactory. It does not tell us whether parents who are financially and psychologically capable of raising a child in a manner we deem proper nevertheless may give up the child for adoption, for what we would consider less than compelling reasons. If parents are morally entitled to give up a child even if the child could have (in some sense) been raised properly by them,[269] our aversion to slavery does not explain why infants are subject only to market-inalienability. There must be another reason why baby-giving is unobjectionable.

The reason, I think, is that we do not fear relinquishment of children unless it is accompanied by market rhetoric.[270] The objection to market rhetoric may be part of a moral prohibition on market treatment of any babies, regardless of whether nonmonetized treatment of other children would remain possible. To the extent that we condemn baby-selling even in the absence of any domino effect, we are saying that this "good" simply should not exist. Conceiving of any child in market rhetoric wrongs personhood. In addition, we fear, based on our assessment of current social norms, that the market value of babies would be decided in ways injurious to their personhood and to the personhood of those who buy and sell on this basis, exacerbating class, race, and gender divisions. To the extent the objection to baby-selling is not (or is not only) to the very idea of this "good" (marketed children), it stems from a fear that the nonmarket version of human beings themselves will become impossible. Conceiving of children in market rhetoric would foster an inferior conception of human flourishing, one that commodifies every personal attri-

---

[269] But perhaps we should prophylactically decline to trust any parents who wished to give a child away for "frivolous" reasons adequately to raise a child if forced to keep her.

[270] Relinquishing a child may be seen as admirable altruism. Some people who give up children for adoption do so with pain, but with the belief that the child will have a better life with someone else who needs and wants her, and that they are contributing immeasurably to the adoptive parents' lives as well as to the child's. Baby-selling might undermine this belief, because if wealth determined who gets a child, we would know that the adoptive parents valued the child as much as a Volvo but not a Mercedes; if an explicit sum of money entered into the decision to give the child up, then we would not as readily place the altruistic interpretation on our own motives. If babies could be seen as incompletely commodified, however, the altruism might coexist with sales. *See supra* pp. 1913–14, 1926.

bute that might be valued by people in other people. In spite of the double bind, our aversion to commodification of babies has a basis strong enough to recommend that market-inalienability be maintained.

The question of surrogate mothering seems more difficult.[271] I shall consider the surrogacy situation in which a couple desiring a child consists of a fertile male and an infertile female.[272] They find a fertile female to become impregnated with the sperm of the would-be father, to carry the fetus to term, to give birth to the child, and to relinquish it to them for adoption. This interaction may be paid, in which case surrogacy becomes a good sold on the market, or unpaid, in which case it remains a gift.

Those who view paid surrogacy as tantamount to permitting the sale of babies point out that a surrogate is paid for the same reasons that an ordinary adoption is commissioned: to conceive, carry, and deliver a baby.[273] Moreover, even if an ordinary adoption is not

---

[271] Surrogacy is often popularly viewed as baby-selling, and the thirteenth amendment is invoked. *See, e.g.*, Holder, *supra* note 268, at 117. The slavery analogy is inadequate for the reasons detailed above.

Surrogacy has engendered a number of different viewpoints. *See, e.g.*, Hollinger, *From Coitus to Commerce: Legal and Social Consequences of Noncoital Reproduction*, 18 U. MICH. J.L. REF. 865, 870 (1985) (arguing that "any legal efforts to prohibit this [baby-making] market from operating would be unwise"); Katz, *Surrogate Motherhood and the Baby-Selling Laws*, 20 COLUM. J.L. & SOC. PROBS. 1, 52–53 (1986) (arguing that surrogate motherhood is "fundamentally different from baby-selling" and could provide "a new solution for infertility"); Krimmel, *supra* note 28, at 35 (maintaining that it is ethically impermissible to bring a child into the world for purposes other than the desire to act as her parents); Mellown, *An Incomplete Picture: The Debate About Surrogate Motherhood*, 8 HARV. WOMEN'S L.J. 23 (1985) (pointing out shortcomings of viewing surrogacy either from the perspective of the liberal ideology of free contract or from the conservative perspective of preserving the traditional family); Note, *Developing a Concept of the Modern "Family": A Proposed Uniform Surrogate Parenthood Act*, 73 GEO. L.J. 1283 (1985) (presenting, with extensive commentary, a statute legalizing surrogacy and regulating the interaction by requiring the participation of doctors, psychologists, and lawyers, limiting compensation of the surrogate to $25,000, prohibiting reduction in compensation if the child is stillborn or impaired, and making specific performance available to both parties); Note, *Reproductive Technology and the Procreative Rights of the Unmarried*, 98 HARV. L. REV. 669, 684–85 (1985) (arguing that the Supreme Court has implicitly recognized a right to procreate and that individuals should not be "arbitrarily deprived of the ability to exercise [that right] through the use of reproductive technology"); Note, *Rumpelstiltskin Revisited: The Inalienable Rights of Surrogate Mothers*, 99 HARV. L. REV. 1936, 1954–55 (1986) (defending the inalienability of abortion rights for surrogate mothers and the alienability of their rights to rear the children once born); *see also* Magisterium of the Catholic Church, Instruction on Respect for Human Life in Its Origin and on the Dignity of Procreation: Replies to Certain Questions of the Day 25 (Feb. 22, 1987) (stating that surrogacy, like artificial insemination by a donor, is "contrary to the unity of marriage and to the dignity of the procreation of the human person").

[272] A full treatment of the issues of surrogacy must also consider embryo transfer, in which the baby is not genetically related to the surrogate, and single men or gay couples who desire to become fathers.

[273] Surrogacy appears even more like a commissioned adoption if what is important to the adopting couple is not primarily the genetic link between father and baby, but rather the opportunity to exercise control over the mother's background and genetic make-up and to monitor her pregnancy. *See, e.g.*, *The Pain of Infertility: One Couple's Choices*, L.A. Times,

commissioned, there seems to be no substantive difference between paying a woman for carrying a child she then delivers to the employers, who have found her through a brokerage mechanism, and paying her for an already "produced" child whose buyer is found through a brokerage mechanism (perhaps called an "adoption agency") after she has paid her own costs of "production." Both are adoptions for which consideration is paid. Others view paid surrogacy as better analogized to prostitution (sale of sexual services) than to baby-selling. They would say that the commodity being sold in the surrogacy interaction is not the baby itself, but rather "womb services."[274]

The different conceptions of the good being sold in paid surrogacy can be related to the primary difference between this interaction and (other) baby-selling: the genetic father is more closely involved in the surrogacy interaction than in a standard adoption. The disagreement about how we might conceive of the "good" reflects a deeper ambiguity about the degree of commodification of mothers and children. If we think that ordinarily a mother paid to relinquish a baby for adoption is selling a baby, but that if she is a surrogate, she is merely selling gestational services, it seems we are assuming that the baby cannot be considered the surrogate's property, so as to become alienable by her, but that her gestational services can be considered property and therefore become alienable. If this conception reflects a decision that the baby cannot be property at all — cannot be objectified — then the decision reflects a lesser level of commodification in rhetoric. But this interpretation is implausible because of our willingness to refer to the ordinary paid adoption as baby-selling.[275] A more plausible interpretation of conceiving of the "good" as gestational services is that this conception reflects an understanding that the baby is already someone else's property — the father's. This characterization of the interaction can be understood as both complete commodification in rhetoric and an expression of gender hierarchy. The would-be father is "producing" a baby of his "own,"[276] but in order to do so he must

---

Mar. 22, 1987, § 6, at 12, col. 1. One adopting father remarked: "We felt, in the case of surrogates, we would be involved from the beginning: conception, monitoring the fetus . . . ." The couple said they "would have adopted had the surrogate option not been available." *Id.*

[274] *See, e.g.*, Hollinger, *supra* note 271, at 893 ("The payments are not to purchase a child, but to compensate for personal services."); *see also* Note, *Baby-Sitting Consideration: Surrogate Mother's Right to "Rent Her Womb" for a Fee*, 18 GONZAGA L. REV. 539, 549 (1983) (arguing that a surrogate mother is not selling her baby, but rather is "provid[ing] a home in her womb for the child of another").

[275] If we were assuming that babies cannot be property, we would more readily envision an ordinary adoption for a price not as baby-selling, but rather as sale of gestational services, or fetal growth support services, followed by the gift of an unmonetized child.

[276] *See, e.g.*, *To Serve "the Best Interest of the Child"*, N.Y. Times, Apr. 1, 1987, § B, at 2, col. 2 ("At birth, the father does not purchase the child. It is his own biological genetically related child. He cannot purchase what is already his."). Indeed, the very label we now give the birth mother reflects the father's ownership: she is a "surrogate" for "his" wife in her role of bearing "his" child.

purchase these "services" as a necessary input. Surrogacy raises the issue of commodification and gender politics in how we understand even the description of the problem. An oppressive understanding of the interaction is the more plausible one: women — their reproductive capacities, attributes, and genes — are fungible in carrying on the male genetic line.[277]

Whether one analogizes paid surrogacy to sale of sexual services or to baby-selling, the underlying concerns are the same. First, there is the possibility of even further oppression of poor or ignorant women, which must be weighed against a possible step toward their liberation through economic gain from a new alienable entitlement — the double bind. Second, there is the possibility that paid surrogacy should be completely prohibited because it expresses an inferior conception of human flourishing. Third, there is the possibility of a domino effect of commodification in rhetoric that leaves us all inferior human beings.

Paid surrogacy involves a potential double bind. The availability of the surrogacy option could create hard choices for poor women. In the worst case, rich women, even those who are not infertile, might employ poor women to bear children for them. It might be degrading for the surrogate to commodify her gestational services or her baby, but she might find this preferable to her other choices in life. But although surrogates have not tended to be rich women, nor middle-class career women, neither have they (so far) seemed to be the poorest women, the ones most caught in the double bind.[278]

Whether surrogacy is paid or unpaid, there may be a transition problem: an ironic self-deception. Acting in ways that current gender ideology characterizes as empowering might actually be disempowering. Surrogates may feel they are fulfilling their womanhood by producing a baby for someone else, although they may actually be reinforcing oppressive gender roles.[279] It is also possible to view would-

---

[277] Biblical "surrogate" interactions may be seen in this way. *See Genesis* 16 (Abraham, Sarah, and Hagar); *Genesis* 30 (Jacob, Rachel, and Bilhah). Perhaps some would see artificial insemination as analogously oppressive to men, but the situations are asymmetrical because of the present gender structure. *See infra* note 285.

[278] *See, e.g., Surrogate Motherhood: A Practice That's Still Undergoing Birth Pangs*, L.A. Times, Mar. 22, 1987, § 6, at 12, col. 2 (citing research finding that "[t]he average surrogate mother is white, attended two years of college, married young and has all the children she and her husband want"). Perhaps allowing surrogacy but not permitting adoption for a price would worsen the double bind for poor women, who are less likely to be chosen as surrogates by the couples who seek this arrangement. To underscore the irony of the double bind, consider the testimony of an adopting mother who fears that surrogacy "can exploit the lower classes and the women of the Third World," and thus finds it "unconscionable" to choose as surrogates women who are poverty-stricken and need the money. *Id.* § 6, at 12, col. 1.

[279] Even if surrogate mothering is subjectively experienced as altruism, the surrogate's self-conception as nurturer, caretaker, and service-giver might be viewed as a kind of gender role-oppression. *See, e.g.*, A. DALLY, INVENTING MOTHERHOOD: THE CONSEQUENCES OF AN IDEAL

be fathers as (perhaps unknowing) oppressors of their own partners. Infertile mothers, believing it to be their duty to raise their partners' genetic children, could be caught in the same kind of false consciousness and relative powerlessness as surrogates who feel called upon to produce children for others. Some women might have conflicts with their partners that they cannot acknowledge, either about raising children under these circumstances instead of adopting unrelated children, or about having children at all. These considerations suggest that to avoid reinforcing gender ideology, both paid and unpaid surrogacy must be prohibited.

Another reason we might choose prohibition of all surrogacy, paid or unpaid, is that allowing surrogacy in our nonideal world would injure the chances of proper personal development for children awaiting adoption. Unlike a mother relinquishing a baby for adoption, the surrogate mother bears a baby only in response to the demand of the would-be parents: their demand is the reason for its being born.[280] There is a danger that unwanted children might remain parentless even if only unpaid surrogacy is allowed, because those seeking children will turn less frequently to adoption.[281] Would-be fathers may strongly prefer adopted children bearing their own genetic codes to adopted children genetically strange to them; perhaps women prefer adopted children bearing their partners' genetic codes. Thus, prohibition of all surrogacy might be grounded on concern for unwanted children and their chances in life.

Perhaps a more visionary reason to consider prohibiting all surrogacy is that the demand for it expresses a limited view of parent-child bonding; in a better view of personal contextuality, bonding should be reconceived. Although allowing surrogacy might be thought to foster ideals of interrelationships between men and their children,[282] it is unclear why we should assume that the ideal of bonding depends especially on genetic connection. Many people who adopt children feel no less bonded to their children than responsible genetic parents;[283] they understand that relational bonds are created in shared

---

(1982); A. RICH, OF WOMAN BORN: MOTHERHOOD AS EXPERIENCE AND INSTITUTION (1976); Hantzis, *supra* note 259, at 696.

[280] This is true whether the surrogate gives or sells the baby or her services (however we wish to characterize the thing transferred). If an adoption is commissioned, the baby would not have been born but for the would-be parents' demand, but probably even if these transactions were permitted there would still be a substantial number of unwanted children also available for adoption.

[281] *See supra* note 273.

[282] People who are sensitive to what men lose by not having the bonds with children traditionally thought characteristic of motherhood might argue that if we hope for "new" men that are more bound up with their children, we should foster progress toward this ideal by assuming a deep and personal bond between men and their genetic offspring. Hence, we might think we should respect and encourage men's desires for surrogacy.

[283] There has been very little study, however, of the emotional aftermath of adoption. *See*

life more than in genetic codes.[284] We might make better progress toward ideals of interpersonal sharing — toward a better view of contextual personhood — by breaking down the notion that children are fathers' — or parents' — genetic property.[285]

In spite of these concerns, attempting to prohibit surrogacy now seems too utopian, because it ignores a transition problem. At present, people seem to believe that they need genetic offspring in order to fulfill themselves; at present, some surrogates believe their actions to be altruistic.[286] To try to create an ideal world all at once would do violence to things people make central to themselves. This problem suggests that surrogacy should not be altogether prohibited.

Concerns about commodification of women and children, however, might counsel permitting only unpaid surrogacy (market-inalienability). Market-inalienability might be grounded in a judgment that commodification of women's reproductive capacity is harmful for the identity aspect of their personhood and in a judgment that the closeness of paid surrogacy to baby-selling harms our self-conception too deeply. There is certainly the danger that women's attributes, such as height, eye color, race, intelligence, and athletic ability, will be monetized. Surrogates with "better" qualities will command higher prices in virtue of those qualities. This monetization commodifies women more broadly than merely with respect to their sexual services or reproductive capacity. Hence, if we wish to avoid the dangers of commodification and, at the same time, recognize that there are some situations in which a surrogate can be understood to be proceeding

---

C. FOOTE, R. LEVY & F. SANDER, CASES AND MATERIALS ON FAMILY LAW 404-24 (3d ed. 1985). As we can recognize from the widespread incidence of child abuse and neglect, not all genetic parents are bonded to their children in any ideal sense.

[284] True, there is usually a deep bond between a baby and the woman who carries it, but it seems to me that this bond too is created by shared life, the physical and emotional interdependence of mother and child, more than by the identity of the genetic material. It will be difficult to study this question unless childbearing by embryo transfer, in which a woman can carry a fetus that is not genetically related to her, becomes widespread.

[285] *See* Smith, *Parenting and Property*, in MOTHERING: ESSAYS IN FEMINIST THEORY 199 (J. Trebilcot ed. 1983). Artificial insemination — and for that matter traditional procreation — poses a similar issue of genetic property. It is just as inappropriate to conceive of parent-child bonding in terms of women's genetic "property" as in terms of men's. But in the context of the present gender structure, the desire to carry on the woman's genetic line is less likely to make men fungible. Moreover, the interests of women and men are asymmetrical because the carrying of the child in the woman's body (whether or not it is hers genetically) is a stronger factor in interrelationships with a child than an abstract genetic relationship.

[286] According to those who arrange surrogacy transactions, some women who have acted as surrogates do report altruistic motivations. *See* N. KEANE & D. BREO, THE SURROGATE MOTHER (1981); *cf.* L.A. Times, *supra* note 273, § 6, at 12, col. 2 (reporting the statement of a psychotherapist for a Beverly Hills surrogacy center that the majority of surrogates say that "they enjoy being pregnant, are attracted by the money . . . , and feel deep sympathy for women who are unable to have children"). *But cf.* note 279.

out of love or altruism and not out of economic necessity or desire for monetary gain,[287] we could prohibit sales but allow surrogates to give their services. We might allow them to accept payment of their reasonable out-of-pocket expenses — a form of market-inalienability similar to that governing ordinary adoption.[288]

Fear of a domino effect might also counsel market-inalienability. At the moment, it does not seem that women's reproductive capabilities are as commodified as their sexuality. Of course, we cannot tell whether this means that reproductive capabilities are more resistant to commodification or whether the trend toward commodification is still at an early stage. Reproductive capacity, however, is not the only thing in danger of commodification. We must also consider the commodification of children. The risk is serious indeed, because, if there is a significant domino effect, commodification of some children means commodification of everyone.[289] Yet, as long as fathers do have an unmonetized attachment to their genes (and as long as their partners tend to share it), even though the attachment may be nonideal, we need not see children born in a paid surrogacy arrangement — and they need not see themselves — as fully commodified. Hence, there may be less reason to fear the domino effect with paid surrogacy than with baby-selling. The most credible fear of a domino effect — one that paid surrogacy does share with commissioned adoption — is that all women's personal attributes will be commodified. The pricing of surrogates' services will not immediately transform the rhetoric in which women conceive of themselves and in which they are conceived, but that is its tendency. This fear, even though remote, seems grave enough to take steps to ensure that paid surrogacy does not become the kind of institution that could permeate our discourse.

Thus, for several reasons market-inalienability seems an attractive solution. But, in choosing this regime, we would have to recognize the danger that the double bind might force simulations of altruism by those who would find living on an expense allowance preferable to their current circumstances.[290] Furthermore, the fact that they are not being paid "full" price exacerbates the double bind and is not really helpful in preventing a domino effect. We would also have to recognize that there would probably not be enough altruistic surro-

---

[287] One such example occurs when a woman bears a child for her childless sister.

[288] To prevent women from benefiting financially from reproductive services, some states have passed criminal statutes prohibiting women who relinquish children for adoption from receiving expenses. Others require a full accounting of fees received. *See* Katz, *supra* note 271, at 8–10, nn. 34–37.

[289] *See supra* pp. 1925–26.

[290] The same worry applies, of course, to baby-selling in jurisdictions in which paying the mother's expenses is allowed. *See supra* note 288.

gates available to alleviate the frustration and suffering of those who desire children genetically related to fathers,[291] if this desire is widespread.

The other possible choice is to create an incomplete commodification similar to the one suggested for sale of sexual services. The problem of surrogacy is more difficult, however, primarily because the interaction produces a new person whose interests must be respected. In such an incomplete commodification, performance of surrogacy agreements by willing parties should be permitted, but women who change their minds should not be forced to perform.[292] The surrogate who changes her mind before birth can choose abortion; at birth, she can decide to keep the baby.[293] Neither should those who hire a surrogate and then change their minds be forced to keep and raise a child they do not want. But if a baby is brought into the world and nobody wants it, the surrogate who intended to relinquish the child

---

[291] In light of the apparent strength of people's desires for fathers' genetic offspring, the ban on profit would also be difficult to enforce. As with adoption, we would see a black market develop in surrogacy.

[292] The issue of whether surrogacy agreements should be specifically performed — whether the mother who changes her mind should nonetheless be forced to hand over the baby — has received the most popular attention recently. *See, e.g.*, *Father of Baby M Granted Custody; Contract Upheld*, N.Y. Times, Apr. 1, 1987, § A, at 1, col. 5. We should not think, however, that we are faced with merely a binary choice: either banning paid surrogacy arrangements or granting specific performance of them. To conceive of surrogacy as a special situation requiring specific performance seems to place undue weight on the supposed genetic interests of would-be fathers in their unique "property," and to undervalue both the personal development of unwanted children they might otherwise adopt (and become bonded to) and the personal identity of women torn between economic need and deep attachment to a baby. *But cf.* Hollinger, *supra* note 271, at 909–19. Hollinger's sensitivity to the effect of surrogacy and other new reproductive strategies on the adoption of children who are not white or middle-class, and hence are less "desirable," and her understanding that the strength of the interest in parenthood need not be as closely tied to genetic parenthood as we have tended to view it, seem at odds with her conclusion that surrogacy contracts should be specifically performed. *See id.* at 909–12.

[293] Of course, we should decide upon a reasonable time limit during which she must make up her mind, for it would be injurious to the child if her life were in limbo for very long. This could be done analogously with statutory waiting periods for adoption to become final after birth. *See, e.g.*, Surrogate Parenting Assocs. v. Kentucky *ex rel* Armstrong, 707 S.W.2d 209, 213 (Ky. 1986) (holding that the five-day waiting period in Kentucky's termination of parental rights statute and consent to adoption statute "take[s] precedence over the parties' contractual commitments, meaning that the surrogate mother is free to change her mind"). We might wish to make the birth mother's decision to keep the child not an absolute right but only a very strong presumption, such as would be used in a custody dispute over a newborn baby in a divorce. In my view, however, adoption is the better analogy: except in very special cases, both surrogates and others who are considering relinquishing children for adoption should be able to decide after birth to keep the child. *See, e.g.*, *id.* (stating that if a surrogate decides to keep her child, "[s]he would be in the same position vis-a-vis the child and the biological father as any other mother with a child born out of wedlock" and that the "parental rights and obligations between the biological father and mother, and the obligations they owe the child," would be those imposed by the statutes applicable to this situation).

should not be forced to keep and raise it.[294] Instead, those who, out of a desire for genetically related offspring, initiated the interaction should bear the responsibility for providing for the child's future in a manner that can respect the child's personhood and not create the impression that children are commodities that can be abandoned as well as alienated.[295]

We should be aware that the case for incomplete commodification is much more uneasy for surrogacy than for prostitution. The potential for commodification of women is deeper, because, as with commissioned adoption, we risk conceiving of all of women's personal attributes in market rhetoric, and because paid surrogacy within the current gender structure may symbolize that women are fungible baby-makers for men whose seed must be carried on. Moreover, as with commissioned adoption, the interaction brings forth a new person who did not choose commodification and whose potential personal identity and contextuality must be respected even if the parties to the interaction fail to do so.

Because the double bind has similar force whether a woman wishes to be a paid surrogate or simply to create a baby for sale on demand, the magnitude of the difference between paid surrogacy and commissioned adoption is largely dependent on the weight we give to the father's genetic link to the baby. If we place enough weight on this distinction, then incomplete commodification for surrogacy, but not for baby-selling, will be justified. But we should be aware, if we choose incomplete commodification for surrogacy, that this choice might seriously weaken the general market-inalienability of babies,

---

[294] Because a pregnancy and a child's life are involved in the surrogacy interaction, rather than just one sexual encounter as with prostitution, "official" recognition of the interaction, with its contribution to commodification, will have to be tolerated, regardless of whether we choose market-inalienability or incomplete commodification. Decisions will have to be made about restitution in case of breach, about payment of the surrogate's expenses, and above all, about care for the child if all parties fail to take responsibility. Even if we choose incomplete commodification, contract remedies should be avoided. Specific performance should be avoided because of the analogy to personal service agreements, and also because we should not conceive of children as unique goods, *see supra* note 292; damage remedies should be avoided because of the obvious "official" commodification involved in setting a dollar value on the loss. It is not my purpose here, however, to try to draft an appropriate statute or guidelines for courts.

[295] The special dangers of commodification in the surrogacy situation should serve to distinguish it from the way we treat children generally. Perhaps a regulatory scheme should require bonding, insurance policies, or annuities for the child in case of death of the adoptive parents or reneging by them. *See* Note, *Developing a Concept of the Modern "Family": A Proposed Uniform Surrogate Parenthood Act*, *supra* note 271, at 1304. *But cf.* Hollinger, *supra* note 271, at 911 n.174 (arguing that financial requirements for surrogate parents are unwarranted because the state does not require that "children generated by coital means be similarly protected"). Perhaps a better scheme (because less oriented to market solutions) could require that alternative adoptive parents at least be sought in advance.

which prohibits commissioned adoptions.[296] If, on balance, incomplete commodification rather than market-inalienability comes to seem right for now, it will appear so for these reasons: because we judge the double bind to suggest that we should not completely foreclose women's choice of paid surrogacy, even though we foreclose commissioned adoptions; because we judge that people's (including women's) strong commitment to maintaining men's genetic lineage will ward off commodification and the domino effect, distinguishing paid surrogacy adequately from commissioned adoptions; and because we judge that that commitment cannot be overridden without harm to central aspects of people's self-conception. If we choose market-inalienability, it will be because we judge the double bind to suggest that poor women will be further disempowered if paid surrogacy becomes a middle-class option, and because we judge that people's commitment to men's genetic lineage is an artifact of gender ideology that can neither save us from commodification nor result in less harm to personhood than its reinforcement would now create. In my view, a form of market-inalienability similar to our regime for ordinary adoption will probably be the better nonideal solution.

VI. CONCLUSION

Market-inalienability is an important normative category for our society. Economic analysis and traditional liberal pluralism have failed to recognize and correctly understand its significance because of the market orientation of their premises. In attempting to free our conceptions from these premises in order to see market-inalienability as an important countercurrent to our market orientation, I have created an archetype, universal commodification, and tried to show how it underlies both economic analysis and more traditional liberal thinking about inalienability. As an archetype, universal commodification is too uncomplicated to describe fully any actual thinker or complex of ideas, but I believe consideration of the archetype and what it entails is a necessary corrective. The rhetoric of commodification has led us into an unreflective use of market characterizations and comparisons for almost everything people may value, and hence into an inferior conception of personhood.

I have created a contrasting archetype, universal noncommodification, to characterize the utopian vision — expressed by Marxists and other social critics of the market order — of a social world free of market relationships and market conceptions. Although this archetype, too, is an oversimplification, I believe it enables us to focus on

---

[296] If paid surrogacy is permitted, it can become a substitute for commissioned adoption. *Cf. supra* note 273.

the transition problem that always lies between us and our utopias. If decommodification of things important to personhood is provisionally the ideal of justice we should strive for, trying to bring it to pass now may sometimes be unjust. In attempting to make the hard choices in which both commodification and decommodification seem harmful — the transition problem of the double bind — we must evaluate each contested commodification in its temporal and social context, and we must learn to see in the commodification issue the same interconnection between rhetoric and reality that we have come to accept between physical reality and our paradigms of thought.

To the extent that we must not assimilate our conception of personhood to the market, market-inalienabilities are justified. But market-inalienabilities are unjust when they are too harmful to personhood in our nonideal world. Incomplete commodification can help us mediate this kind of injustice. To see the world of exchange as shot through with incomplete commodification can also show us that inalienability is not the anomaly that economics and more traditional liberalism conceive it to be. This perspective can also help us begin to decommodify things important to personhood — like work and housing — that are now wrongly conceived of in market rhetoric.

Market-inalienability ultimately rests on our best conception of human flourishing, which must evolve as we continue to learn and debate. Likewise, market-inalienabilities must evolve as we continue to learn and debate; there is no magic formula that will delineate them with utter certainty, or once and for all. In our debate, there is no such thing as two radically different normative discourses reaching the "same" result. The terms of our debate will matter to who we are.

This page intentionally left blank

OPTING OUT OF THE LEGAL SYSTEM: EXTRALEGAL CONTRACTUAL RELATIONS IN THE DIAMOND INDUSTRY

*LISA BERNSTEIN**

I. INTRODUCTION

BUSINESS disputes arise in all industries, and the diamond industry is no exception. Unlike the situation in many other industries, however, diamond industry disputes are resolved not through the courts and not by the application of legal rules announced and enforced by the state. The diamond industry has systematically rejected state-created law. In its place, the sophisticated traders who dominate the industry have developed an elaborate, internal set of rules, complete with distinctive institutions and sanctions, to handle disputes among industry members. This article explores the reasons that this system of private governance has developed and endured within the diamond trade.

Section II provides a brief overview of the diamond industry. It sketches the workings of the international diamond cartel and discusses diamond production and valuation. Section III describes the organization of the market for rough and polished diamonds, paying special attention to the role of trading clubs (bourses). It focuses on the terms and structure of transactions and details the workings of the bourse's private arbitration system that keeps all judgments secret as long as they are promptly paid.

Section IV is the core of this article. Section IV*A* briefly considers why diamond dealers need to make executory contracts; it then explains that the diamond market also operates as an implicit loan market. Section IV*B* compares the cost of entering into legally unenforceable (extralegal) agreements to the cost of entering into legally enforceable contracts. It

* Associate Professor, Boston University School of Law. I would like to thank Steven Shavell, Louis Kaplow, Lucian Bebchuk, and David Charny. The John M. Olin Foundation provided funding for this project.

[*Journal of Legal Studies*, vol. XXI (January 1992)]
 0047-2530/92/2101-0002$01.50

concludes that the transaction costs of entering into legally enforceable agreements cannot explain diamantaires' preference for extralegal agreements and suggests that the norm of "secrecy" that pervades the industry is at least a partial explanation for diamond dealers' preference for privately enforced agreements.

Section IV*C* considers the characteristics of public law that make contracts enforced through litigation an unattractive option. Its primary focus is on the way courts calculate expectation damages. It argues that, if commercial transactions in the diamond industry were governed solely by legally enforceable contracts under which the promisee could recover expectation damages in the event of breach, the market would be characterized by frequent, inefficient breach of contract. It attributes this inefficiency to the uncertainty of recovery, the way courts calculate damages, the time it takes to obtain a judgment, and the fact that many diamantaires do not have ready access to capital markets.

Although many of the shortcomings in the American legal system that make litigation unattractive to diamond dealers are also present in most commercial contexts, the diamond industry is unique in its ability to create and, more important, to enforce its own system of private law. Sections IV*D* and *E* focus on the ways the industry's organization facilitates the creation of strong reputational bonds, which the bourse's arbitration system in turn uses to enforce its judgments. They examine two types of reputation-bond-based extralegal contractual regimes: the homogeneous group regime that is generally associated with repeat transactions among members of small geographically concentrated and ethnically homogeneous groups, and the information-intermediary regime in which technology links markets and secures the rapid and low-cost dissemination of information about reputation. Although the industry is currently moving from a homogeneous group to an information-intermediary-based regime, it has succeeded, at least for the time being, in creating an overarching system that captures the advantages of both regimes.

Section IV*F* explores some of the efficiency implications of reputation bonds, relative to those created by court-imposed expectation damages. Section IV*G* discusses the substantive and procedural reasons why arbitration is preferred to litigation. Finally, Sections IV*H* and *I* assess the aggregate efficiency of the system and the importance of reputation bonds in the market as a whole. In sum, the analysis presented in Section IV suggests that, while the damage rules adopted by the industry may lead to some instances of inefficient breach, the system's overall success is due, in large part, to its ability to quickly resolve disputes and enforce judgments—results that cannot be obtained through the legal system.

Section V uses a model of arbitration and settlement to explain why

most intraindustry disputes are resolved cooperatively and without recourse to a third-party arbiter. Section VI considers the changes in the industry that have led to the gradual introduction of legally enforceable written contracts in certain types of diamond transactions. It also discusses the increasing influence of civil law on the terms of diamond transactions and the resolution of disputes. Section VII concludes that the diamond industry provides strong support for the hypothesis that extralegal norms trump legal rules in a given market only where market participants find that keeping to the industry norms advances their own self-interest. The private regime must be Pareto superior to the established legal regime in order to survive.[1]

II. An Overview of Diamond Production and Valuation

The market for rough and polished gem-quality diamonds is best understood in the context of the chain of production and distribution that begins in a pit mine and ends up in a retail jeweler's window.[2] Rough diamonds are found primarily in Africa, Australia, and the Soviet Union; they are not notably rare. At present, 80–85 percent of the world's supply of rough diamonds is controlled by the DeBeers Cartel. The cartel distributes its supply of rough diamonds through four brokers. The brokers then sell presorted boxes of diamonds to some 150–200 dealers, known as sight holders,[3] during ten viewing sessions, or sights, held in London each year. Most U.S. sight holders are members of the New York Diamond Dealers Club (DDC). At a sight, a dealer is given a box of diamonds and informed of its price. This price is nonnegotiable. If the dealer decides not to purchase his box, he will not be invited to subsequent sights. Consequently, a sight holder will rarely decline to purchase his box,

[1] A similar thesis is advanced by Robert C. Ellickson, A Hypothesis of Wealth Maximizing Norms: Evidence from the Whaling Industry, 5 J. L. Econ. & Org. 83, 84 (1989), where he explores the "hypothesis that when people are situated in a close knit group, they will tend to develop for the ordinary run of problems norms that are wealth maximizing." For a broader and more theoretical articulation of this thesis, see generally, John Gray, Hayek on Liberty (2d ed. 1986), discussing Hayek's theory of spontaneous order and the "competitive selection of rules and practices," particularly, "Hayek's assertion that the evolution of culture may itself be fruitfully investigated in terms of the competition between different traditions or practices, with a natural selection among them occurring which is at least partly to be explained by their relative efficiency as bearers or embodiments of knowledge."

[2] The markup from mine to consumer is estimated to be between 200 and 400 percent.

[3] There are three kinds of sight holders: large manufacturers who cut and polish the stones themselves, midsize rough dealers who resell the contents of their boxes to select manufacturers, and brokers who deal in industrial diamonds. Polished stones are sold to wholesalers and marketed through brokers, both of whom then sell to retail establishments.

although he will sometimes negotiate in advance to sell it unopened to another dealer on a cost-plus-profit basis.

The cartel insists that the diamonds be paid for in full within seven days of the sight. Consequently, for most sight holders, particularly those who cut and polish the rough themselves, access to credit is essential—it takes three–four months from the sight date for a manufacturer to sort, cut, polish, and sell the contents of his box. Sight holders, however, rarely have difficulty securing financing. In the diamond industry, having a sight is considered a near guarantee of financial success. The cartel actively monitors the decisions and activities of sight holders; if a sight holder continues to play by the cartel's "rules," he is rewarded with a more profitable selection of stones. Consequently, because most monitoring costs are shifted to the cartel, sight holders generally have access to bank capital.

Diamond valuation is a subjective process. The value of a rough diamond depends on the value of the polished stones that can be manufactured from it. Since no two diamantaires will cut a stone the same way, the value added in the manufacturing process varies widely. Consequently, when dealers value a piece of rough differently, that difference, on which profitability turns, often reflects a real difference in the value of the polished stones they will be able to cut from it.

In contrast, when dealers value a polished stone differently, most of that difference will be due to their differing estimates of market demand and to their differential skill in detecting flaws in stones. In recent years, however, the skill factor has become less important. Although older dealers continue to maintain that even polished diamonds cannot be objectively graded and valued, in the late 1970s, the Gemological Institute of America began to issue diamond grading certificates whose widespread use made it possible for dealers with little gem expertise to enter the market, resulting in increased competition. By creating standardized ways of describing polished stones, grading certificates have facilitated the flow of price information. A private diamantaire now publishes a weekly price list with a wide international circulation.[4]

As a consequence of the standardization of grading and the availability of price lists, the market for polished diamonds has become more com-

[4] Unlike a closing quotation on a typical commodities exchange, the prices recorded in the *Rapaport Diamond Report* are not actual transaction prices. Rather, they are the Rapaport Corporation's subjective calculation of the "high asking" price, generally 15–30 percent above the actual transaction price, for various sizes and grades of polished stones. One explanation for the markup is that it enables retailers to quote list prices to consumers who think they are getting a bargain when they buy below it.

petitive and prices have dropped. This has reduced the profit margin of manufacturers since retailers now have more reliable information about what wholesalers and manufacturers have paid for stones. Manufacturers find themselves squeezed between the price of rough fixed by the cartel and the competitive prices in the polished market.

III. The Market for Rough and Polished Diamonds

A. The Trading Club as a Commodities and Information Exchange

The largest and most important trading club ("bourse") in the United States is the New York Diamond Dealers Club. Its membership is comprised of sight holders, manufacturers, wholesalers, and brokers. Club membership gives a dealer prestige and an important economic advantage. In the diamond industry, access to a steady supply of goods is essential to the operation of a profitable brokerage or manufacturing business. Although it is possible to buy stones on the "open market," a dealer who does not have access to the trading clubs—essential links in the worldwide diamond distribution network—will be at a competitive disadvantage. Approximately 80 percent of the rough diamonds coming into the United States pass through the hands of a DDC member, as do 15–20 percent of the polished stones. In addition, 20–50 percent of the transactions conducted by or on behalf of foreign or out of town dealers are concluded in the club.[5]

The New York DDC currently has 2,000 members; in most years there is a waiting list for admission. Although requirements for membership are strict, the main constraint on membership is space, not the inability of dealers to meet the membership requirements.[6] As a condition of mem-

[5] Despite the strict limits on the number of members it accepts, the DDC tries to attract out-of-town dealers (and nonmembers) to its trading hall. Before being admitted to the trading hall, out-of-town dealers must be introduced by a member in good standing who agrees to assume "full financial responsibility (guarantee) for the out of town dealer's acts and liabilities, incurred while on the premises of the DDC." Diamond Dealers Club Bylaws (hereinafter DDC Bylaws), Art. 17 § 2a (1980). Consequently, nonmembers who want access to supply find it advantageous to maintain a reputation for scrupulous honesty with club members. The out-of-town dealer must also be approved by the board of directors, pay a fee determined by the board, and agree to adhere to all of the club's bylaws, including the obligation to arbitrate all disputes. In return for his sponsorship, a member who introduces an out-of-town dealer is entitled to collect a commission of 1 percent on every transaction the out-of-town dealer consummates.

[6] To be considered for membership a dealer must (1) have been in the industry for at least two years, (2) comply with all requests for information put to him by the board of directors, and (3) have his picture posted in the club for ten days so that members have the opportunity to state reasons that he should not be accepted. New members are put on

bership, a dealer must sign an agreement to submit all disputes arising from the diamond business between himself and another member to the club's arbitration system.[7] The agreement to arbitrate is binding. Unless the club opts not to hear the case, the member may not seek redress of his grievances in court. If he does so, he will be fined or expelled from the club. Furthermore, since the agreement to arbitrate is binding, the court will not hear the case.

Most large and important dealers are members of the club, but they do not usually conduct their business in the club's trading hall. In the diamond industry, where profitability depends largely on a dealer's network of contacts, secrecy is valued; large-scale transactions tend to be consummated in private offices. In addition, because properly valuing a stone depends on the ability to detect minor flaws and color variations, buyers prefer to examine large stones in familiar light. Furthermore, for security reasons, many dealers do not want it known that they have valuable stones in their possession. Larger dealers will, however, come to the club's trading hall to get a feel for market prices. As one dealer explained, a visit to the club enables him to "keep a finger on the pulse of the business." Although a price list is available for certain classes of polished stones, the bourse's trading floor is the only place to obtain a feel for the market price of rough diamonds: standardized price information is unavailable. Unlike other commodities exchanges, the DDC itself does not record either actual transactions prices or the volume of transactions.[8]

Smaller dealers, brokers, and foreigners do most of their trading in the club. For them, club membership provides a secure trading place at a modest cost with additional informational benefits. In general, the reputations of smaller dealers are less well established. Club membership en-

---

probation for a period of two years during which "the Board of Directors reserves the right to terminate such membership at any time within this period for any reason." *Id.* at Art. 3 § 8. New members are charged a $5,000 initiation fee, and annual dues are $1,000.

Although corporations may designate individuals to become members of the club and to trade on their behalf, these individuals do not enjoy limited liability as they would under the civil law. See Diamond Dealers Club Arbitration Bylaws (hereinafter DDC Arbitration Bylaws), Art. 12 § 25 (1987). The corporation or partnership is also considered liable and bound by the members' agreement to submit all disputes to the DDC arbitration system. DDC Bylaws, Art. 3 § 2b.

The traditional view of diamond trading as a family business is reflected in the membership bylaws: more lenient rules govern the admission of sons, daughters, sons-in-law, and daughters-in-law. For example, *id.* at Art. 3 § 2a, provides that widows of members are automatically accepted and do not have to pay an initiation fee. Similarly, the "wife of an incapacitated member may be accorded entry into the Club at the sole discretion of the Board of Directors *until her husband becomes active*." *Id.* at Art. 3 § 3b.

[7] See *id.* at Art. 12 § 1c.

[8] The Federal Trade Commission estimates that 700–800 dealers use the club each day.

ables them to signal that they are trustworthy and, conversely, gives them the assurance that all the dealers in the trading hall have fulfilled the requirements for club membership, an important non-transaction-specific piece of information.

The bourse is an information exchange as much as it is a commodities exchange. As one author put it, "the bourse grapevine is the best in the world. It has been going for years and moves with the efficiency of a satellite communications network. . . . Bourses are the fountainhead of this information and from them it is passed out along the tentacles that stretch around the world."[9] The bourse facilitates the transmission of information about dealers' reputations[10] and, at least with respect to members, serves both a reputation-signaling and a reputation-monitoring function.[11]

The New York DDC is a member of the World Federation of Diamond Bourses (WFDB), an umbrella organization composed of the world's twenty diamond bourses.[12] A dealer who is a member of any one bourse in the world federation is automatically allowed to trade at all member bourses. Each bourse has similar trade rules, and, like the individual bourses, the WFDB has an arbitration system to resolve differences between its members. As a condition of membership in the federation, each bourse is required to enforce the arbitration judgments of other member bourses to the extent permitted by the law of the country in which it operates.[13]

B. The Standard Transactional Paradigm

In the diamond industry, a handshake accompanied by the words *mazel u'broche* creates a binding agreement. Section One of the Trade Rules

[9] V. Berquem, Bourses More than a Place to Sell, Jewellery News Asia (August 1988).

[10] For example, the DDC's bulletin boards carry letters from dealers who feel they have been victimized by baseless gossip. These letters contain rebuttals and frequently include strong language condemning the integrity of dealers who spread baseless rumors. Sometimes, in addition to being posted, such letters are distributed in the trading hall or on Forty-seventh Street itself.

[11] As stated in its bylaws, the purposes of the club are, among other things, "to inculcate just and equitable principles in trade, to eliminate abuses and unfair trade practices relative thereto or affecting the same, to diffuse accurate and reliable information concerning the matters relating thereto, [and] to produce uniformity in the conduct of business ethics." DDC Bylaws, Art. 2.

[12] The DeBeers Cartel does not control the WFDB. One of the main reasons the WFDB was formed was to enable the bourses to bargain more effectively with the cartel. See, generally, Albert Lubin, Diamond Dealers Club: A Fifty Year History (1982).

[13] For more information on the rules of the World Federation, see World Federation of Diamond Bourses: Bye-Laws and Inner Rules (unpublished report, World Federation of Diamond Bourses, November 15, 1988).

provides: "[a]ny oral offer is binding among dealers, when agreement is expressed by the accepted words 'Mazel and Broche' or any other words expressing the words of accord."[14] While older dealers continue to adhere to this tradition in a steadfast manner, younger dealers who tend to be less well acquainted with their trading partners and more worried about the prospect of misunderstanding or breach often memorialize the key terms in writing, despite the fact that such a writing is not required for formation of a binding agreement.

The most common transactional paradigm is known as "open cachet."[15] When a buyer makes an offer to a seller or a broker, the stone is put in an envelope which is then folded and sealed in a precise way. The terms and conditions of the offer are placed on the envelope as is the date. The buyer then signs the parcel across the seal. Unless otherwise specified, this offer is considered binding on the offerer until one o'clock the next day. The seller may accept at any time during this period by contacting the buyer and saying "mazel and broche." If, however, the seller either rejects the offer or makes a counteroffer during this period, his option to accept the buyer's original offer is canceled.

There are sound business reasons for the use of written terms on a cachet parcel. If the buyer who made the offer and created the binding option contact cannot be reached by a seller who wants to accept the offer within the proscribed period of time, the seller, assuming that he is a member of the club, is entitled to "place his acceptance of the offer, in writing, on the same wrapper and have the time of his acceptance certified by a member of the Board of Directors of the Diamond Dealers Club."[16] Without this formality and its attendant trade rule, a buyer who regretted making an offer could simply refuse to see visitors or take phone calls until the cachet period had elapsed. A buyer's reachability is difficult and costly to monitor given the variety of plausible excuses that could be invented. Thus, a more explicit form of contract is used to overcome the weakness of the reputation bond. Finally, in open cachet transactions

[14] DDC Bylaws, Art. 18 § 1.

[15] Another type of cachet used less frequently than open cachet is known as a "Zee'ch," or "search" cachet. In a Zee'ch transaction, the seller seals the stone and signs the parcel. This signals his agreement not to show the stone to anybody else for a period of twenty-four hours. A Zee'ch seal does not give the buyer an option to purchase the sealed stone at a particular price. Rather, it gives him an exclusive right to resume negotiations for the stone at a specified time in the future. It is common for buyers to shop around by putting a variety of stones under Zee'ch. This practice makes sense in the market for rough stones where no standardized price information is available; it makes comparison shopping easier, which facilitates competitive pricing.

[16] DDC Bylaws, Art. 18 § 3.

there is generally no such thing as opportunistic breach by a seller. Although it is improper for the seller to show the stone to another buyer while it is still under cachet, the wrapper is signed across the seal to discourage this and make its violation known to the buyer. Nonetheless, if the seller receives a higher offer on the stone from someone who viewed it prior to consummation of the cachet, the trade rules permit him to terminate the cachet by contacting the original cachet holder. It is customary, however, for him to tell the original buyer of the new offer and to give him the opportunity to match it. This leads to a miniauction with the stone being sold to the highest bidder.

Although the cachet is formally an agreement between a buyer and a seller, its most important function in the market is to regulate the relationship between a seller and his broker. If no cachet were used and the buyer offered five hundred dollars per karat for the stone, the broker might tell the seller that the buyer offered four hundred dollars per karat. If the seller accepted, the broker would pocket the difference. This is not the type of dishonest behavior that could be easily monitored and enforced through reputation bonds since detection and a determination of the precise circumstances would be difficult.

When a deal is physically concluded on the floor of the DDC, a document akin to an integrated writing is frequently, but not always, produced. After the parties have either made an oral agreement or gone through the formalities of cachet—that is, at a stage in the transaction where the parties already consider themselves bound—they take the goods to be weighed by a club employee who issues them an official weight slip. The slip is then signed by the person who gave the stone to the club official, most commonly, though not exclusively, the buyer, and the terms of payment and the price are added. If the sale was concluded in accordance with the rule of open cachet, the cachet parcel is included in the bag with the stone and the official weight slip. Only one copy of the slip exists, and it is retained by the seller. If a dispute later occurs, the club's dispute resolution bodies consider the slip to be definitive evidence of both the stone's weight and the existence of the transaction.

In some instances, stones are traded not on the floor of the DDC but in private offices. In this case, a standard bill of sale is drawn up when the deal is concluded and before the buyer leaves with the stones. Frequently, however, it is sent by the seller after the buyer has left with the stone but before he has paid in full. Dealers explain that when they really trust the person they are trading with they do not, at the time of "contracting," attach any real importance to this writing. Traditionally, the bill of sale has been viewed as a mere formality used primarily for accounting purposes.

Even if either the weight slip or the bill of sale satisfied the requirements of the New York Statute of Frauds, a suit could not be brought in either New York State Court or federal court since the club membership agreement requires that all disputes between club members be arbitrated, and this agreement has been upheld as binding.

C. The Club's Private Arbitration System

Around 150 disputes per year are submitted to the DDC's arbitration system. Of these, an estimated 85 percent are settled during the mandatory prearbitration conciliation procedure. Although there has been a slight increase in the number of arbitrations in recent years, this is attributed primarily to the increase in club membership and not to a deterioration of trade ethics.

The DDC's procedural rules clearly reflect the industry's preference for the voluntary resolution of disputes. The bylaws are structured to give the parties control over the dispute resolution process and to create financial incentives to settle. For example, prior to an arbitration hearing, the parties are required to participate in a conciliation proceeding, and "whenever an adjustment by conciliation is consummated, the chairman of [the three-person conciliation] panel may refund the arbitration fee or any part of the same."[17]

An important feature of the arbitration system is the secrecy of the proceedings. The arbitrators are not required to make findings of fact and do not produce written decisions explaining their reasoning. As long as judgments are complied with, the fact of the arbitration as well as its outcome are officially kept secret.

Procedural Aspects of Arbitration. There are two dispute resolution bodies in the DDC, the Floor Committee and the Board of Arbitrators; both are composed of club members elected for two-year-terms. Before a dispute is referred to arbitration, the Floor Committee must find that a material issue of fact exists. The standard used is similar to the familiar standard for granting summary judgment.

The Floor Committee has the authority to exclude a member from the trading hall for up to twenty days and/or impose a fine of up to $1,000 when the member "fails to meet his commercial obligations to another member and no material issue of fact is involved or a member causes a disturbance or conducts himself in the clubrooms in a manner unbecoming a member of the club."[18] A decision of the Floor Committee may be

[17] DDC Arbitration Bylaws, Art. 12 § 8.

[18] *Id.* at Art. 8 § 7B1.

appealed by filing a written request and paying the $100 appeal fee. Unless the panel finds that a material issue of fact exists and recommends that the case be referred to arbitration, the decision of the appeal panel is final. Neither the Floor Committee nor the appeal panel are required to make any findings of fact.

Any member of the DDC who has a claim "arising out of or related to the diamond business"[19] against another member has the right to file a written complaint against the member who must then submit to DDC adjudication. At the time he files the complaint, the plaintiff must pay a small arbitration fee,[20] but at the conclusion of the case the panel "shall decide which of the litigants shall pay the arbitration fee and the expenses which were necessarily incurred, and . . . may refund the arbitration fee or any part of it."[21] Arbitrators are required to render their decision within ten days of the hearing.

Arbitration awards can be appealed if notice of appeal is filed with the board of directors within ten days of the parties' receipt of the judgment. The appellant must pay a fee three times the original arbitration fee and "deposit cash or sufficient security to cover the amount of the judgment."[22] The appeals board is composed of five arbitrators who did not hear the original case, and it too is "under no obligation to specify any findings of fact which are reversed or modified nor set forth any new findings of fact."[23]

The decisions of the arbitration board can be appealed to New York State court under New York law, but arbitration awards can only be vacated for procedural irregularities, such as an arbitrator engaging in an ex parte communication or a failure to allow the parties to be represented by counsel.[24] The substantive rule of decision is not reviewed.

[19] *Id.* at Art. 12 § 1a.

[20] *Id.* at Art. 12 § 2.

[21] *Id.*

[22] *Id.* at Art. 12 § 15.

[23] *Id.* at Art. 12 § 17.

[24] See, for example, Goldfinger v. Lisker, 508 N.Y.S.2d 159 (on a motion to confirm a DDC arbitration award the court granted a cross motion to vacate the award on the grounds that, while the DDC Bylaws do authorize arbitrators to investigate the facts, ex parte communications with arbitrators are not thereby sanctioned). In addition, New York law requires that the arbitration process be free from the appearance of bias. See, for example, Rabinowitz v. Olewski, 100 A.D.2d 539; 473 N.Y.2d 232 (2d Dept 1984) (where the court ordered a stay of DDC arbitration and directed that the case be heard by an independent arbitrator after a letter surfaced in the club accusing the plaintiff of being sympathetic to the Palestine Liberation Organization; since it was clear that a substantial injustice might result were the case heard by the predominantly Jewish DDC and there was the "appearance of impropriety and specter of bias among the DDC").

Although the DDC arbitration system is operated primarily for the benefit of club members, nonmembers who have a dispute with members often request that the club hear their case. In most instances, the board will grant their request as long as the member consents and both parties sign an agreement to arbitrate. There are a number of reasons why nonmembers might request that the DDC arbitrate a dispute with a member. First, if the nonmember knows he is in the wrong, yet the parties are unable to agree on a settlement, then having a neutral third party assess a penalty should enable him to minimize the reputation cost of his breach since arbitration awards are kept secret if the judgment is paid promptly. Although the arbitration's results sometimes become known through gossip, as long as the individual is not frequently involved in such controversies, the damage to his reputation is likely to be contained. Second, if the nonmember thinks he is in the right, arbitration is preferable to litigation because it is cheaper, faster, and subjects the member to unqiue pressures to pay promptly. Although club members are not obligated to submit disputes with nonmembers to arbitration, they will often agree to do so in order to avoid the transaction and reputation costs of going to court.

Substantive Aspects of Arbitration. The DDC Board of Arbitrators does not apply the New York law of contract and damages, rather it resolves disputes on the basis of trade customs and usages. Many of these are set forth with particularity in the club's bylaws, and others simply are generally known and accepted. Although at first glance diamond transactions appear to be simple buy-sell agreements, complicated controversies often arise, particularly in the sale of polished stones. In general, disputes fall into three main classes: those that have explicit remedies prescribed in the trade rules;[25] those that have no explicit remedies prescribed but are common enough that they are dealt with consistently according to widely known customs; and those complex disputes that the arbitrators either decline to hear or decide in accordance with rules of decision and damage measures that neither party can predict ex ante.

The dispute resolution system in the diamond industry shows some sensitivity to concerns of institutional competence. Under its bylaws, the club has the right to refuse to arbitrate a claim when it does not arise out of the diamond business, or "(1) involves complicated statutory rights; (2) is 'forum nonconveniens' in that it is burdensome or inconvenient to handle the claim in the Club; (3) involves nonmembers; (4) has been

[25] See DDC Bylaws, Trade Rules II: Customs and Usage. For the rules governing transactions in certificate stones, see DDC Arbitration Bylaws, Trade Rules Regarding Certificate Stones.

conciliated, mediated, arbitrated or litigated outside the Club and/or the parties have sought remedies elsewhere; (5) is not in the ordinary course of commercial dealings."[26] When the club refuses to hear a case, the parties are permitted to seek remedies outside the club.

In complex cases that are neither explicitly covered by the trade rules nor dealt with according to established custom, it is difficult to determine what substantive rules of decision are applied. Arbitrators explain that they decide complex cases on the basis of trade custom and usage, a little common sense, some Jewish law, and, last, common-law legal principles. There are *no* general rules of damages. When calculating damages, the arbitrators look at the stone, consider the circumstances, and apply their business experience. Many dealers feel that the arbitrators have redistributive instincts; they cite the unpredictability of the decisions as well as the arbitrators' tendency to "split the difference" as an important motivation to settle their disputes on their own. This may be a reason why, while 150 arbitration complaints are filed each year, only thirty to forty go to judgment. The arbitrators announce their judgment, but they neither make findings of fact nor explain their reasoning. The absence of explicit findings of fact and written opinions is a precaution to prevent people from complaining, rightly or wrongly, that the arbitrators were biased, unfair, or relied on evidence that lacked probative value. The arbitration board is like a jury black box. Diamond dealers eschew arbitration for many of the same reasons that businessmen in general are wary of jury trials, primarily the uncertainty of the outcome.

A person who is found to have breached an agreement or engaged in unethical conduct is sometimes ordered to pay punitive damages or a fine in the form of a donation to charity in addition to compensating the other party for his loss. Thus, unlike court awards that, while unpredictable, are at least bounded by expectation damages, arbitration awards have a completely uncertain component. In one case, a dealer falsely accused another dealer of stealing a stone. The accuser subsequently remembered where he had put the stone and apologized to the other dealer. As the incident had become widely known throughout the club, however, the wrongly accused dealer brought an arbitration action against the owner of the stone for impugning his good name. The board ordered the man to make a full public apology and a fifty thousand dollar donation to a Jewish charity.

[26] *Id.* at 12 § 1b. See, for example, Finker v. The Diamond Registry, 469 F. Supp. 674 (S.D.N.Y. 1979) (where the DDC agreed to decide issues concerning the ownership of goods held on memorandum (consignment) but "refused to involve itself in the dispute concerning the trademark registration and alleged infringement").

Enforcing Arbitration Judgments. The DDC Bylaws provide that "[a]ll decisions of arbitration panels including floor committee arbitrations which are not complied with within 10 working days, together with the picture of the non-complying member, shall be posted in a conspicuous place in the Club rooms."[27] This information is communicated to all bourses in the world federation. As a condition of membership in the federation, each bourse agrees to enforce the judgments of all member bourses. Since most diamond dealers frequently transact in foreign bourses, this reciprocity of enforcement greatly increases the penalty for failing to voluntarily comply with an arbitration judgment.

The arbitration board can also suspend or expel a member for failing to pay a judgment or failing to pay his diamond-related creditors without making special arrangements through the club's private bankruptcy system.[28] Unlike the arbitration system, which operates in place of a public trial, the DDC's bankruptcy rules and procedures do not supplant civil bankruptcy law; they provide instead a parallel set of rules that are mandatory for club members: "[a]ny settlements made outside of the jurisdiction of the Club do not absolve the debtor member's liability for suspension purposes."[29] There is no such thing as "discharge" under the private bankruptcy rules: "All debtors must make provisions for the payment of one hundred percent (100%) of his/her debt";[30] debt is rescheduled on the basis of the dealer's ability to pay.

After the club has been notified of a member's bankruptcy, the member is required to "turn over in escrow to the Diamond Dealers Club, Inc. his assets of any kind for distribution to his creditors,"[31] and a creditors committee is formed to effect the distribution. While bankruptcy proceedings are taking place, the debtor is not allowed to enter the club room unless given explicit permission to do so by the club committee.[32] Similarly, "where the debtor has requested a settlement with his creditors for any sum less than one hundred percent (100%), and has not complied

[27] DDC Arbitration Bylaws, Art. 12 § 26.

[28] See, generally, DDC Bylaws, Art. 20.

[29] *Id.*, Art. 20 § 18. In addition, a member is automatically suspended from the club for "filing a petition in bankruptcy or any involuntary petition in bankruptcy, [or] making an assignment for the benefit of creditors." DDC Arbitration Bylaws, Art. 7 § 1.

[30] DDC Bylaws, Art. 20 § 18.

[31] *Id.* at Art. 20 § 2a.

[32] See, for example, Matter of Marcus [MVAIC], 29 Misc.2d. 573, In Matter of Paul Verstandig v. Diamond Dealers Club, Inc. 23 A.D.2d 547 (1965) (upholding "the Club's action in suspending petitioner as a member for the breach of the Debtor-Creditor General Rules of the Club").

with the action required of him as set forth in this article,"[33] he may be suspended or expelled from the club and his name is circulated to all of the bourses in the world federation and posted on their bulletin boards. The bankruptcy rules are strictly enforced since the industry depends on credit reliability.

After conclusion of bankruptcy proceedings, "[a] majority of the Board of Directors may reinstate any suspended member should they feel s/he has conducted her/himself as a bona fide debtor and has made provisions for the payment of one hundred percent (100%) of his/her debt."[34] Formerly bankrupt members who comply with the club's bankruptcy rules are sometimes readmitted under this provision.

In general, the Board of Arbitrators uses suspension more frequently than expulsion to secure compliance with its decisions. Expulsion presents a classic end-game problem. The expelled member may feel like he has nothing to loose by challening the club—he can try to upset the board's decision in court, file a private antitrust suit, or sue in tort for interference with business relations. The bylaws, however, provide that a member who was suspended or expelled may be readmitted after two years on the same terms as a new member. Although this provision appears to be a partial solution to the end-game problem, due to the long waiting list of those already qualified for club membership and the subjectivity of the admissions process, dealers are not routinely readmitted under this provision. Furthermore, even if the admissions committee voted to readmit a dealer, his ability to avoid being shunned would depend on the original reason for his expulsion. The bylaw provision was probably included to enable the club to avoid charges of intentional interference with business relations.[35]

Under New York law, binding arbitration awards can be confirmed in civil court. If this is done, the judgment has the same force and effect as an initial court award. In practice, however, it is rarely necessary for a party to a DDC arbitration to seek confirmation of a judgment. While arbitration awards are officially kept secret, a confirmation proceeding in court would quickly become public knowledge. Thus, the dealer against whom the judgment was entered would suffer severe damage to his reputation. Furthermore, if a member refuses to pay a judgment and the party

[33] DDC Bylaws, Art. 20 § 11a.

[34] *Id.* at Art. 20 § 19.

[35] In the wake of an antitrust suit brought against the club in 1951, challenging the club's practice of refusing to deal with Germans after World World II, an article was added to the bylaws that cautions members not to engage in any behavior that can be construed as being in restraint of trade.

who prevailed finds it necessary to obtain a court enforcement order, the DDC bylaws require the losing party to pay an additional 15 percent of the award to cover his opponent's legal expenses. Another enforcement mechanism sometimes invoked by the arbitrators is a proceeding in Jewish rabbinical courts against the party who refuses to comply. Because these courts have the authority to ban an individual from participation in the Jewish community, this is a powerful threat against Orthodox members of the diamond industry.

IV. An Economic Analysis of the Extralegal Contractual Regime

A. The Reasons That Executory Agreements Are Needed

In order to understand contractual relations in the diamond industry, it is important to briefly consider why executory agreements—contracts—are used at all. For many transactions, simultaneous exchange is advantageous. It reduces the riskiness of the transaction, decreases transaction costs by eliminating costly and time consuming negotiations over payment terms, eliminates the need for going through the formalities of cachet, and, most important, enables dealers to trade with people about whose reputation they have little information. Simultaneous exchange is facilitated by the presence of a major diamond-financing bank in the same building as the DDC. In addition, the seven other banks that extend credit to New York dealers are located nearby.

Although simultaneous exchange frequently occurs, particularly in small-scale transactions, it is neither possible nor beneficial in many instances. There is a great need for credit in the diamond industry. As explained above, even the largest sight holders need credit to finance the purchase of their boxes of rough. Similarly, non–sight holders also acquire most of their stones on a cycle that follows, but lags behind, the schedule of sights. They therefore need credit to enable them to purchase enough stones to keep their cutters working until the next sight. Access to credit is also essential in the market for polished stones. Because polished stone sales are highly seasonal, with 30–40 percent occurring in November and December, access to credit is needed to avoid a cash shortfall.

After the diamond crash of the early 1980s,[36] banks became more reluc-

[36] As a result of a confluence of factors, the diamond industry suffered a severe crash in the late 1980s. Throughout the mid- to late 1970s, the Israeli banks began extending low interest rate loans to rough dealers at the direction of the Israeli government, who wanted to expand the diamond cutting industry in Tel Aviv. The only collateral required was the

tant to finance diamond dealers, particularly small dealers and non–sight holder manufacturers. As a consequence, bargaining over the term of payment became an important and contentious stage in contract negotiation. The most common terms are immediate cash payment, thirty-day terms, and sixty-day terms. The thirty- and sixty-day periods correspond roughly to the time it takes to manufacture a stone. This varies depending on the cut, the stone, and the skill of the manufacturer. The close correlation between cutting time and the length of the payment terms suggests that sellers generally finance most, if not all, of the buyer's (manufacturer's) cash gap.

The market for rough and polished diamonds functions not only as a commodities market but also as an implicit capital market.[37] One possible explanation for the extension of credit by sellers is that sellers typically have better and less expensive access to outside capital than most buyers. Many of the important sellers are also DeBeers sight holders. The fact that a dealer is a sight holder sends a signal to the bank that he is a good credit risk. Banks prefer to lend to sight holders because they need not incur the large cost of valuing gems that they would have to bear if they lent to non–sight holders whose inventories are in a constant state of flux. Lenders can offer lower interest rates to sight holders because they can have greater confidence when they make loans and most monitoring costs are shifted to the cartel.

Unlike banks, sight holders are industry insiders; they have good information about individual dealers' reputations and transact with the same people on a repeat basis over a long period of time. It is thus cheaper for sight holders to monitor dealers' reputations and credit worthiness than it is for banks. Consequently, it is likely that sight holders can offer terms

---

diamonds actually purchased, which were stored in the banks' vaults. When the world economy entered a recession in 1980, dealers found it difficult to resell their diamonds and, as the price began to fall, they defaulted on loans. The banks found themselves with a 1.5 billion dollar stockpile of diamonds, more than the DeBeer's Cartel could afford to repurchase to maintain the price. Although an agreement between the cartel, the Israeli government, and the Israeli banks was finally reached that prevented the entire stockpile from being immediately released into the market, enough were resold to drastically lower prices.

At around the same time the "investment diamonds" scheme that had been developed in the late 1970s, whereby telephone salesmen sold sealed packages of "investment" grade diamonds to investors along with a promise to repurchase them at a later date, also went bust when the companies failed and purchasers tried to cut their losses by selling the diamonds on the open market, further depressing prices.

[37] In addition to the sale of goods on credit, the practice of giving goods on consignment (memorandum), which is common in transactions between wholesalers and retailers, is a way of effecting an implicit loan.

(and an implicit interest rate) that a buyer would prefer to simultaneous *exchange* financed through a short-term bank loan.[38]

The economics of the diamond industry suggest that there must be a way for dealers to make and enforce executory contracts. Sections IV*B* through *E* discuss why the diamond industry opts for extralegal agreements over legally enforceable contracts and considers the two ways that these extralegal agreements are enforced. Sections IV*F* through *I* consider the efficiency implications of these arrangements and discuss the substantive and procedural reasons that arbitration is preferred to litigation.

B. The Choice between Extralegal Agreements and Legally Enforceable Contracts

One line of analysis used to explain market transactors' choice between legally enforceable contracts and extralegal contracts focuses on the transaction costs of negotiating and drafting legally enforceable agreements. It is not clear, however, a priori that these costs are necessarily higher than those incurred in the formation of an extralegal contract consummated with a handshake. Because the ability of the promisee to enforce an extralegal contract depends on the posting of a reputation bond by the promisor, each of the parties must bear the "information cost" of determining whether the other party is trustworthy before negotiation over the terms of the agreement even begins.[39] This cost may be substantial and will depend, at least in part, on the size, structure, and terms of the proposed transaction as well as on the likelihood that the parties will have occasion to deal with one another again in the near future.

[38] The Merchants Bank of New York, however, is attempting to create a market niche for itself by creating a special group of gem experts who become involved in the day-to-day operations of the industry (thereby gaining access to intraindustry reputation information). The bank then extends short-term loans to non–sight holder dealers. The bank's policy, however, is new and it is too early to assess its success.

[39] Although in the typical diamond transaction the buyer takes possession of the stone and promises to pay the seller at some time in the future, the buyer must still obtain information about the seller's reputation. Using lasers and chemical processes, diamonds can be treated to artificially enhance color and disguise flaws. Small flaws and differences in color dramatically affect the value of a stone. Many of these "treatments," however, cannot be detected without sophisticated equipment. Although in theory buyers could have every stone evaluated by a gemological laboratory to determine whether or not it had been altered, this would be prohibitively time consuming and expensive. Nevertheless, if a dealer purchases a "treated stone" and sells it to someone else who discovers the stone's treatment, he can be taken to the arbitration panel for failing to disclose the treatment. The panel must then decide whether the dealer knew or reasonably should have known of the stone's treatment. The reputation of the person he purchased the stone from is an important factor considered by the arbitrators. See also note 64 *infra*.

In general, the magnitude of precontract transaction costs incurred in the formation of extralegal contracts will depend on how common such contracts are in the relevant market. In the diamond industry, extralegal contracts are the dominant contractual paradigm. Consequently, the industry is organized to minimize the cost of obtaining information about dealers' reputations.

In addition to the bourse system, which rapidly transmits reputation information, the precontract transaction cost of entering into an extralegal contract is reduced by the use of brokers. Brokers are able to gather information about individuals' reputations for trustworthiness at a lower effective cost than individual buyers and sellers because a broker's investment is less transaction specific. When a buyer and a seller invest in acquiring information about their respective reputations only to find that the buyer needs a particular size stone that the seller does not have, the parties have lost part of their investment. While the information acquired may be useful to them in the future, its value diminishes over time as its accuracy decreases. In contrast, a broker who has this information can shop around immediately for new trading partners for either party. As the geographical dispersion of the industry increases, brokers are becoming more important.[40]

In a market where enforcement depends on social ostracism or reputational damage, the formation of an extralegal contract depends on information about reputation. In addition, it requires adherence to enough formalities to alert other members of the relevant group that an agreement has taken place. In the diamond industry this function is served by customs such as handshakes, cachets, weight slips, and bills of sale that are able to effectively serve the channeling, cautionary, and evidentiary functions of formality while imposing minimal additional cost.[41]

[40] Despite their informational advantage, there are a variety of factors that limit brokers' role in the market. Information about "trustworthiness," unlike consumer credit information, is difficult to communicate in objective terms. What is ethical behavior to a thirty-year-old dealer, may be an abhorrent business practice to a sixty-year-old dealer. (One Israeli dealer explained that within the bourse there are small trading groups whose members trade primarily among themselves. The groups are defined by their standards of what constitutes fair and ethical trading.) In addition, even when they participate in brokered transactions, the individual buyer and seller still have to acquire some information about the broker's judgment and reputation. This task is cheaper, however, since it is less transaction specific. Once a dealer determines that a broker has good judgment, he has access to many other dealers whose reputations he need not inquire into directly. Even so, the brokerage fee of 1 percent of the price may be prohibitively high, particularly on low-profit-margin transactions.

[41] See Lon Fuller, Consideration and Form, 41 Colum. L. Rev. 799 (1941) (distinguishing between: (1) the channeling function of contract in which "form offers a legal framework into which the party may fit his actions. . . . it offers channels for the legally effective

The diamond industry's preference for an extralegal contractual regime cannot be explained by the transactions costs incurred in preserving an agreement in an integrated writing. Although the industry is organized to minimize the costs of using extralegal agreements, given the widespread use of weight slips, invoices, and bills of sale, the additional transaction costs of using legally enforceable standard-form contracts would not be significant. These agreements could be drafted to approach a complete contingent-state contract because most of the events that might disrupt a transaction are well known within the industry and are subject to well-established customs and usages within the trade. Nonetheless, the use of such agreements is not observed.

In the diamond industry, even if fully specified legally enforceable contracts were widely used and could be inexpensively drafted, dealers would still incur many of the precontract transaction costs of entering into extralegal agreements. In general, as the cost of enforcing a contract in court increases relative to the expected benefit, even fully specified legally enforceable contracts contain an implicit and increasingly large extralegal component. This is also true when the expected value of the court-awarded remedy is insufficient to fully compensate the promisee. In the typical diamond transaction, litigation costs would be high relative to the amount that could be recovered, and the promisee would almost always be undercompensated under standard damage remedies.[42] Therefore, even if legally enforceable contracts were used, diamond dealers would still need the benefit of the reputation bond posted in the formation of an extralegal agreement. And, consequently, dealers would still have to incur the transaction costs of inquiring into their trading partner's reputation and conforming to industry custom.

In addition, a legally enforceable agreement, no matter how cheap to draft and easy to enforce, is not usually considered to be a positive "good" in the diamond industry. Secrecy is highly valued,[43] and whoever makes public the workings of the business will suffer a loss in the value of his reputation, even if he is merely defending himself against a meritless lawsuit. Consequently, individual traders, fearful of litigation that might reveal trade practices, prefer to conclude transactions using agreements

---

expression of intention"; (2) the "cautionary or deterrent function by acting as a check against inconsiderate action"; and (3) the evidentiary function "of providing evidence of the existence and purpose of the contracts in the case of controversy").

[42] See text at Section IV*C infra*.

[43] From the perspective of insiders, secrecy raises high barriers to entry that reduce potential competition. Secrecy also helps ward off unwanted government regulation of the market.

that are enforceable only in the bourse's arbitration tribunals, where the existence and outcome of a dispute are kept secret as long as the judgment is paid promptly. Given the well-established institutional premium on secrecy, parties are rarely willing to pay the reputational price of violating that norm simply to gain access to the courts. Historically, preserving the secrecy norm is one of the primary reasons that the industry uses extralegal agreements rather than legally enforceable contracts.

Although the secrecy norm's strength has been diminishing in recent years, the industry's preference for extralegal contracts remains strong. In general, parties are more likely to opt for extralegal contracts whenever there are costs or factors that the courts are systematically unwilling to recognize or take into account in setting damages (either for doctrinal or public policy reasons) but that ex ante both parties perceive as being important.[44] The same is also true when the courts refuse to apply a rule of decision preferred by the parties or, in interpreting agreements, refuse to do so in light of the prevailing custom. In sum, extralegal contracts are more likely to become an industry norm in situations where traditional contract remedies are likely to lead to inefficiently high levels of breach of contract and the market is organized in a way that makes other methods of enforcing these agreements possible. In the diamond industry, both of these conditions are met.

C. The Shortcomings in the American Legal System and the Common Law of Damages That Make Extralegal Contracts Desirable to Diamond Dealers

If commercial transactions in the industry were governed solely by explicit, legally enforceable contracts under which the promisee could recover expectation damages in the event of breach, the market would be characterized by frequent inefficient breach of contract. The sources of this inefficiency are the uncertainty of recovery, the way courts calculate damages, the length of time it takes to obtain a judgment, and, in some instances, the fact that many diamantaires do not have ready access to capital markets. In most settings, expectation damages, as enforced through the courts, do not achieve their stated theoretical objective of placing the promisee in the same position that he would have been in if the breach had never occurred; they neither make the promisee whole ex post, nor give the promisor sufficient incentive to perform the promise ex ante.

[44] A similar argument is advanced by Richard A. Epstein, Beyond Forseeability: Consequential Damages in the Law of Contract, 18 J. Legal Stud. 112–13 (1989).

In practice, courts are reluctant to award compensation for lost profit since in most instances it is considered speculative. In a diamond transaction, when a seller fails to deliver a stone, lost profit is extraordinarily difficult to calculate since it is highly idiosyncratic.[45] A dealer's profit on a rough stone depends intimately on his network of contacts, his skill as a cutter, and his ability to choose a cut for which market demand is high. The same is true of polished stones, but to a lesser degree. Similarly, when a buyer breaches a promise to pay money, it is difficult, if not impossible, to determine the profit the promisee would have made subsequent to the breach had he been able to invest the money he was owed—the value of business opportunities forgone is inherently speculative. The longer it takes to obtain a judgment, which in New York court can take up to three years, the greater will be the uncompensated loss suffered by the promissee when his ability to enter into subsequent transactions is impaired due to lack of capital.

In calculating expectation damages, courts award interest to compensate the promisee for doing without the money during the pendency of the controversy. Interest will fully compensate the promisee only if the unavailability of funds did not affect his ability to enter into subsequent transactions, that is, if the promisee had access to credit on reasonable terms during the relevant time period. The typical diamond dealer does not have ready access to capital markets or excess cash on hand. For example, in a transaction between two non–sight holders, if the promisee is not paid, it is unlikely during the pendency of the dispute that he will be able to either borrow money or obtain access to the implicit capital market at the predispute implicit interest rate.[46] If the amount owed is large, it is quite possible that he will have to suspend operations until he is paid. In the New York diamond market, which specializes in the largest

[45] Even the expert diamantaires who sit on the DDC's arbitration panel have difficulty accurately valuing lost profit and business opportunities forgone. They are allowed to award punitive damages, however, and usually err on the side of generously compensating the promisee for alleged lost profit. They render their decisions quickly so as to minimize the number of business opportunities the promisee will have to forgo when he is not paid. See text at Section IV*G infra*.

[46] A dealer's ability to obtain credit through the industry's implicit capital market will also be affected by the existence of a public dispute. Until a decision is rendered, the judgment enforced, and the dealer absolved of wrongdoing, other dealers will be either unwilling to sell to him on credit or will charge him a higher implicit interest rate on each transaction to compensate for the perceived increase in the risk of nonpayment and the depletion of his cash reserves. Thus, even if a court were to adjust its award of expectation damages by the correct interest rate, the promisee would still be undercompensated; until the controversy is resolved, the promisee will have to pay the higher implicit interest rate in every subsequent transaction, while the court will award him interest only on the amount of the original debt. Thus, the shortcomings in the expectation remedy are particularly acute in an implicit capital market.

and highest quality goods, this is often the case, particularly for midsize dealers who operate on a tight cash flow margin. Having a portion of his capital tied up for three years while a lawsuit progresses through the New York Courts could cause a dealer extensive financial harm that would not be taken into acocunt in the final calculation of damages. In addition, when the promisor's default causes the promisee to breach other contracts, the promisee will suffer long-term damage to his reputation for which he will not be compensated under standard damage measures.

One possible way to contract around some of these difficulties would be to include a liquidated damages clause. The validity of a liquidated damage clause, however, is often uncertain because of the elusive distinction between valid clauses that are "genuine preestimates" of the anticipated damages and those that are void as "penalties."[47] In a diamond transaction, it would be particularly difficult to draft a liquidated damages clause that a court would view as a "good faith" attempt to preestimate damages. Often, at the time of contracting, the parties themselves are unable to accurately preestimate damages since the actual harm suffered by the promisee in the event of breach depends largely on business decisions made after entering into the contract. For example, even if at the time of contracting nonpayment would neither have bankrupted the promisee nor caused him to default on other obligations, if he subsequently made a large financial commitment in reliance on being paid and then was not, he might suffer tremendous financial and reputational harm, particularly if forced to go to court to obtain a judgment. Since at the time of contracting the magnitude of this harm could not have been predicted, liquidated damages clauses designed to compensate the promisee for this type of harm would run a serious risk of being invalidated as penalties. Furthermore, even if a valid clause could be drafted, the cost of negotiating its terms would greatly increase precontract transaction costs, thus depriving the clause of much of its utility.[48]

[47] The validity of a liquidated damages clause is governed by UCC § 2-718(1), which provides that: "[d]amages for breach by either party may be liquidated in the agreement but only at an amount which is reasonable in the light of the anticipated or actual harm caused by the breach. . . . A term fixing unreasonably large liquidated damages is void as a penalty." In Equitable Lumber Corp. v. IPA Land Dev. Corp., 38 N.Y.2d 516, 381 N.Y.S.2d 459, 344 N.E.2d 391 (1976) the court held that, even if a liquidated damages clause "satisfy[s] the test set forth in the first part of § 2-718(1), a liquidated damages provision may nonetheless be invalidated under the last sentence of this section if it is so unreasonably large that it serves as a penalty rather than a 'good faith' attempt to preestimate damages."

[48] Another possibility would be to include a clause making the promisor liable for all consequential damages suffered by the promisee. Because consequential damages can be enormous, are highly unpredictable, and will depend largely on actions taken by the promisee subsequent to the agreement, it is unlikely a businessman would agree to them.

Although the divergence between the expected cost of breach to the promisor and the actual loss suffered by the promisee is likely to be particularly large in the diamond industry, this divergence is present to some extent in every commercial transaction. A suit for breach of contract is a way for the promisee to control the damage he suffers; it does not make him whole. In a now-famous study, Macaulay found that, even among businessmen who use legally enforceable contracts, when unforeseeable contingencies arose, there was a tendency to renegotiate contracts and settle disputes rather than resort to litigation.[49] Because expectation damages never fully compensate the promisee, all business contracts have an implicit, extralegal term that captures the value of the promisors' reputation.

D. Reputation Bonds as a Way of Enforcing Extralegal Contractual Commitments

In practice, a significant portion of most commercial contracts are backed, at least in part, by a reputation bond. What is unique about the diamond industry is not the importance of trust and reputation in commercial transactions, but rather the extent to which the industry is able to use reputation/social bonds at a cost low enough to create a system of private law enabling most transactions to be consummated and most contracts enforced completely outside the legal system.

Types of Reputation Bonds. The typical diamond transaction involves the posting of a reputation bond equal to the present value of the profit on future transactions that will not take place if the promisor breaches a contract, less his ability to cover. In practice, the value of an individual's reputation is a function of the degree to which he possesses those attributes that other dealers consider important in business relationships—such as honesty and a record of prompt payment of debt. In the diamond industry, reputation bonds are, in practice, the sole enforcement mechanism in transactions between dealers who are not members of a bourse. In transactions between bourse members, agreements can also be enforced in a proceeding before the bourse's board of arbitrators, which has the authority to award any measure of damages it finds appropriate and suspend or expel members for noncompliance with its judgments. Reputation bonds, however, are the primary reason that the arbitration tribunal's decisions are obeyed; they are essential to the bourse's ability

[49] Stewart Macaulay, Non-contractual Relations in Business: A Preliminary Study, 28 Am. Soc. Rev. 55 (1963).

to enforce its judgments. The main function of both the club and its arbitration system is to enhance the functioning of reputation bonds.[50]

Transactions between members of the same trading community also involve the posting of a "psychic/social" bond. There are two types of social bonds. Primary social bonds are similar to reputation bonds in that they have a market value. When a primary social bond is sacrificed, a dealer's ability to communicate information about his reputation and obtain information about business opportunities is diminished. In contrast, secondary social bonds may have a value to the individual on a personal level, but their loss often will not have a direct economic effect on the promisor. When a secondary social bond is sacrificed, a dealer may experience, "loss of opportunities for important or pleasurable associations with others, loss of self-esteem, feelings of guilt, or an unfulfilled desire to think of himself as trustworthy and competent."[51] Although secondary social bonds are becoming less important in the diamond industry, vestiges of their former importance remain. The Diamond Dealers Club still functions like an old-fashioned mutual-aid society. It provides kosher restaurants for its members. A Jewish health organization provides emergency medical services, and social committees are organized by neighborhood to visit sick members and their families. There is a synagogue on the premises, and contributions to a benevolent fund are required. Group

[50] In considering the theory of reputation bonds, it is important to keep in mind that the club's ability to enforce its arbitration judgments, whether through fines, suspension, or expulsion, depends on its ability to harness the force of a reputation bond and that the DDC can only enforce its judgments if noncompliance results in forfeiture of a type of reputation bond that is recognized and given value by market forces.

In the early 1980s the DDC Board of Directors exercised their authority to expel a member from the club for making public statements that tended to cast the industry in a negative light. They expelled Martin Rapaport for saying to the press, "diamonds, ethics, Feh! If the devil himself showed up they would sell to him." The real reason the club wanted to expel Rapaport, however, was that they were opposed to his price list. See note 4 *supra*. They also brought an antitrust suit against him for price-fixing and asked a Jewish court to issue an injunction barring him from any further participation in the Jewish community until he ceased publishing the list. The attention generated by the suit led the Federal Trade Commission (FTC) to initiate an investigation to determine if the club itself was in restraint of trade. Although the FTC instituted a full-scale investigation, it was later dropped. Rapaport challenged his expulsion in court but ultimately settled with the club on undisclosed terms and was readmitted as a member. Today Rapaport has a strong base of support at the club: he is a member of the board of arbitrators, and his price list is an accepted fixture in the international diamond trade. Rapaport was not expelled for breaching contracts or failing to meet his commercial obligations; consequently the club was unable to use its power to exclude him from the industry. The norms of the diamond industry only work when they capture information that the market values.

[51] David Charny, Nonlegal Sanctions in Commercial Relationships, 104 Harv. L. Rev. 393 (1990).

discounts on packaged family vacations are also available so that members' families can travel together during the month that the bourse is closed. In addition, the board of directors has the discretion to make charitable contributions of up to 5 percent of the organization's total annual income.

The Theory of Reputation Bonds. Reputation bonds are generally assumed to be effective only within geographically concentrated, homogeneous groups who deal with each other in repeated transactions over the long run. Charny has noted, however, that a reputation-bond-based, extralegal contractual regime will function even in large scale markets so long as "technology . . . such as computers used to monitor credit worthiness, or mass media used in advertising, [make it possible to] convey information cheaply to a large group of transactors . . . [that is] when a thick set of informational intermediaries" exists.[52] The diamond industry is currently in transition; it is moving from a homogeneous-group-based, extralegal contractual regime to one that relies increasingly on information technology.

The Homogeneous Group Regime. In a given market, geographical concentration, ethnic homogeneity, and repeat dealing may be necessary preconditions to the emergence of a contractual regime based on reputation bonds. As the diamond industry illustrates, however, these conditions are not required for the maintenance of such a system, particularly when the system has already demonstrated itself to be preferable (Pareto preferred) to the established legal regime.

In general, homogeneous-group-based, extralegal contractual regimes are more likely to arise when "preexisting or gradually evolving social relationships provide a basis for nonlegal [extralegal] commitment[s] without large additional investments in developing a bond . . . [since they are] incrementally less costly as nonlegal [extralegal] sanctions when they are parasitic on background habits or understandings built into the culture in which these bonds are formed."[53] Because the diamond industry has long been dominated by Orthodox Jews,[54] it was able to take advantage

[52] *Id.* at 419.

[53] *Id.* at 423–24.

[54] Jews have been involved in the diamond industry since the Middle Ages. The original reasons for their involvement were largely fortuitous: Jews happened to live in major cities on the diamond trade route. In two of these cities, Amsterdam and Antwerp, laws relating to Jewish employment were quite liberal and the governments allowed them to freely enter the diamond-cutting trade. The concentration of Jews in the industry accelerated in 1492 when Spain expelled its Jews and large numbers then fled to Amsterdam and Antwerp. For a brief history of the Jewish involvement in the diamond trade, see Abe Shainberg, Jews, Diamonds and History, 100 Israel Diamonds 46 (1987). The continued Jewish involvement in the industry has also been attributed to the fact that, due to the periodic outbreaks of

of the existence of these conditions. In the past, Jews formed a cohesive, geographically concentrated social group in the countries in which they lived. Jewish law provided detailed substantive rules of commercial behavior, and the Jewish community provided an array of extralegal dispute resolution institutions. The parallels between Jewish law and the modern organization of the diamond industry are striking. For example, under Jewish law, a Jew is forbidden to voluntarily go into the courts of non-Jews to resolve commercial disputes with another Jew. Should he do so, he is to be ridiculed and shamed.[55] Jewish law also provides rules governing the making of oral contracts and lays down rules for conducting commercial arbitration.[56] In the diamond industry, Jewish law provided a code of commercial fair dealing that gradually adapted to meet the industry's changing needs; yet, even as the force of religious law broke down, the system remained strong.

The stability observed in homogeneous markets can endure even if there are occasional breakdowns in the mechanism of extralegal enforce-

---

anti-Semitism, Jews sought out forms of wealth that could be easily concealed, transported, and liquidated during difficult times.

[55] See Menachem Elon, ed., The Principles of Jewish Law 20–21 (1974) ("A striking expression of the religious and national character of Jewish law is to be found in the prohibition on litigation in the gentile courts . . . to which the halakhic scholars and communal leaders attached the utmost importance. . . . any person transgressing the prohibition was deemed to have reviled and blasphemed and rebelled against the Torah").

[56] There are many similarities between the DDC Bylaws and Jewish law. Jewish law requires a three man arbitration panel. In complex cases, these Jewish arbiters "generally based their decisions on communal enactments . . . trade usages, . . . appraisal, justice, and equity . . . and at times even upon a particular branch of a foreign legal system." *Id.* at 23. Jewish law also reflects a preference for the voluntary resolution of disputes. Jewish arbitrators were given the authority to attempt to bring about conciliation (compromise) between parties prior to rendering their decision. Jewish arbitrators were also required to schedule hearings and render decisions promptly. Just as the DDC arbitrators are not required to produce written opinions of their decisions, "according to talmudic halakhah [Jewish law], a party may require the regular court to submit written reasons for its judgments, but an arbitral body is not obligated to do so, even upon request." *Id.* at 569. Sometimes, however, "it is considered desirable to make known the reason for a judgment," and this is in fact the practice in the Israeli bourse, which publishes important statements of principle that are used to decide novel questions. The similarity in the terms of the substantive law is also striking. According to Jewish law "any custom adopted by the local merchants as a mode of acquisition is valid . . . since it fulfills the principle that the purpose of the kinyan [any formal act of acquisition] is to bring about the decision of the parties to conclude the transaction. . . . some authorities even regard a handshake as the equivalent of an oath." *Id.* at 209. In addition, under Jewish law, "the decision of the parties to conclude a sale is finalized by the performance of one of the appropriate acts of kinyan ("acquistion") by one of the parties—generally the purchaser—that the other parties have expressed their agreement that this be done. Ownership thereupon passes, regardless of the question of possession, since possession sometimes accompanies the passing of ownership and sometimes not. If the consideration for the sale is monetary payment, pay the purchase price and it becomes a debt for which he is liable." *Id.* at 211.

ment. Sugden developed a model of exchange that demonstrates how, under certain conditions, a market norm that normally results in cooperation can be a stable, though not unique, equilibrium—even when there appear to be incentives for individuals to be free riders and transactors occasionally make mistakes (breach unintentionally).[57] The game is an adaptation of the classic prisoner's dilemma model in which the following conditions hold: the benefit to player 1 of player 2 refraining from defecting b must be greater than the cost to player 1 of refraining from defecting himself c; the same must also be true for player 2; ϕ, the probability that a subsequent round will be played, must be greater than b/c, since, if this condition did not hold, the expected gain from defection will be greater than any gain from alternative strategies.

In the context of the diamond market, these conditions seem to hold. The probability that the transactors will have occasion to deal with one another in the future, ϕ, is quite high. In addition, many aspects of the industry suggest that the condition that $b > c$ will hold. For example, a diamond dealer generally operates on a slim cash flow margin and has trouble getting access to capital. He routinely makes business decisions in reliance on receiving payment on a particular date. If he is not paid, the harm he suffers can be far greater than loss of the amount of money he is owed. Nonreceipt of payment might force him to breach a contract with another dealer, which will in turn damage his reputation. It might force him into insolvency and result in suspension from the club. Overall, he might do better forgoing the benefit of opportunistic breaches and being able to rely on receipt of payments owed.

In essence, the game relies on the familiar strategy of tit-for-tat, in which one player (say an established dealer) agrees to comply with the rules of the game until the other side violates them but will punish that player by defecting from the cooperative solution if the other player has done so in the previous round. In general, the established dealer adopts a strategy that promises cooperation to those who cooperate and punishment to those who defect. That strategy is rendered enforceable by the large number of established dealers willing to play in accordance with the rules of the game. These dealers take an initial position of cooperation and signal their future behavior by bourse membership. So long as the occasional deviations from the basic rules are met by effective punishment, the game can continue indefinitely even though there is less than perfect compliance.

Although in theory the game may be unstable since there remains a

[57] Robert Sugden, The Economics of Rights, Cooperation, and Welfare (1986).

risk of defection where the gains derived from breaching a single contract exceed the net reputational loss, in practice this is unlikely to happen since the largest stones are usually sold at public auction rather than in the club or private offices—not only to obtain a reliable market price, but also to minimize the prospect of opportunistic breach.

E. The Shift toward an Information Technology–based Contractual Regime

Although diamond dealing was one a predominantly Jewish profession, this is no longer true. Today, the World Federation of Diamond Bourses has twenty member bourses, many of which are located in Asia. The industry is increasingly turning to technology to solve the problems created by ethnic diversity and geographical separation. This shift is opposed by older dealers accustomed to dealing primarily with friends and longstanding business acquaintances. As younger dealers are elected to executive positions in their bourses, however, the WFDB is considering a number of far-reaching proposals: setting up an international computer data base with reports of arbitration judgments from all member bourses in an attempt to foster international uniformity in trade customs and a rule requiring that every bourse be equipped with a fax machine for rapidly transmitting credit information. Also under consideration, although staunchly opposed by many dealers, is the creation of an international computer data base describing goods available for sale worldwide.

As the diamond industry has become less ethically homogeneous and more geographically dispersed, the WFDB had encouraged the creation of new bourses. The world federation, in this instance backed by the Central Selling Organization and its tremendous market power, has been able to induce dealers in many countries to set up bourses and pay their share of the monitoring costs needed to maintain the extralegal system. These organizations make it clear to new entrants, who are primarily manufacturers of small stones, that their ability to secure a steady flow of rough diamonds for their cutting centers is intimately linked to their willingness to play by the established rules—to organize bourses, set up arbitration systems, and submit claims filed against them to the Arbitration Board of the World Federation.

Intrabourse monitoring is an effective way of ensuring the continued viability of a system based on reputation and trust. A bourse's ability to attract business depends largely on the aggregate reputation of its members for trustworthiness and fair dealing, and a bourse's economic viability depends, in large part, on its ability to attract foreign dealers to its trading halls. For example, at the New York Diamond Dealers Club,

25–50 percent of the transactions that take place on the premises are by or on behalf of foreign entities or dealers. Diamond trade journals contain many articles about the reputations of various bourses, with particularly heavy coverage being given to new ones. If dealers in these new trading centers want to compete in the international market, they are forced to incur the cost of setting up a bourse and monitoring the reputations of its members.

Intrabourse reputation monitoring, induced by competition between bourses, is likely to be cheaper than increased monitoring by an umbrella organization such as the world federation. Within each bourse, there is a measure of social and ethnic homogeneity. Consequently, intrabourse monitoring can take advantage of preexisting social relationships and therefore be achieved at a lower cost than regulation by an outside body that cannot take advantage of these preexisting relationships.

In general, the world federation's drive to create new bourses has succeeded in combating an additional problem associated with markets based on social networks among homogeneous groups, namely, that "these markets may become unstable because of free-riding potential, as outlying transactors may adopt the customs of the markets without bearing the costs of membership."[58] Although it may be true that, in the long run, "markets based upon social networks are unlikely to sustain themselves in the face of alternative markets based on sophisticated and potentially more extensive information systems,"[59] the diamond industry is currently in a state of transition; it has succeeded, at least for the time being, in creating a system that is designed to capture the benefits of both monitoring by small social groups (individual bourses) and monitoring achieved through information intermediaries (institutions such as the world federation and brokers).

Although trade practices and customs have remained largely unaffected by the shift from a homogeneous-group-based contractual regime toward one that is based increasingly on information technology, the change could radically affect the economic structure of the industry. In a homogeneous-group-based contractual regime, developing a reputation for trustworthiness and fair dealing takes time since reputation information is communicated solely by world of mouth and depends largely on personal contacts. This results in high barriers to entry.[60] In contrast,

[58] David Charny, Implicit Contracts 50 (unpublished manuscript, Harvard Law School, Law and Economics Workshop 1990).

[59] *Id.*

[60] In addition, new entrants, particularly in the manufacturing sector, would also face higher capital requirements than existing market participants since their access to the implicit loan market will be limited until they establish a reputation for trustworthiness. See text at Section IV*A supra*.

information-technology-based regimes lower barriers to entry by reducing an individual's cost of informing others about his reputation either directly or through information intermediaries such as Dun & Bradstreet, Standard & Poors, and, in the retail jewelry business, the credit ratings of the Jewelers Board of Trade.

It might be argued that an outsider with no established reputation could overcome reputation-related barriers to entry by offering to transact using legally enforceable contracts. If extralegal contracts are rationally preferred, however, a promisor offering a written agreement would have to offer a much higher price to compensate the promisee for the risk and imperfections of litigation—not only the actual cost and uncertainty of the litigation, but also the reputational damage of being involved in a court suit at all. More important, over a certain range of transaction values, a legally enforceable agreement is not of great value to a party. Even with larger transactions, the expected value of a legally enforceable contract in the absence of information about the other party's reputation might be less than the expected value of a legally unenforceable agreement with a person with a reputation for honesty and fair dealing. Although, as the size of the transaction increases, the benefit of a legally enforceable contract increases relative to the transaction costs of litigation, the amount of capital that is tied up is greater, which in turn increases the opportunity cost of doing without the capital during the pendency of the litigation.

F. Reputation Bonds and Economic Efficiency

The use of reputation bonds to enforce contracts is sometimes said to be inefficient because there is no correlation between the damage suffered by the promisee and the cost of breach to the promisor. Because the cost of breach to the promisor is generally assumed to be large, reputation bonds are said to induce an inefficiently high level of contractual performance. The most common type of executory agreement in the diamond industry, however, is exchange of goods today for a promise to pay X dollars on a future date. Consequently, the most common type of breach is nonpayment. On the day payment is due and the buyer has to make the decision to perform or breach, the seller's expectancy is known with certainty; it is X dollars. Since only money is at stake, and it is of equal value to both parties, performance is always indicated; the extent of a payment obligation cannot be made to turn on either party's "need" for the money. Thus, even a legal rule that led to no breach of contract would be efficient in the context of these transactions. This is, in fact, close to what is observed in the market; breach of contract is rare. A rule that

leads to no breach of contract has additional benefits in the diamond industry where sellers routinely rely on buyer's promises to pay.

In the market for polished stones, even when transactions take the form of an exchange of executory promises, there is no such thing as efficient breach. Although they are somewhat more difficult to value objectively, and cannot quite be bought and sold on a spot market, polished diamonds are much like any other commodity. Rough diamonds, in contrast, are mere inputs (along with capital, technology, and labor) into the production of polished diamonds. Consequently, an efficient market for rough stones is one in which each rough stone finds its way to its highest valued use. This outcome corresponds to the manufactuer who is willing to pay the most for it since a manufacturer's ability to estimate the value of the polished stones he can make from a piece of rough is critical to his ability to earn a profit.

It might seem that, if seller (S) promises to sell a stone to manufacturer 1 ($M1$) for one hundred dollars and manufacturer 2 ($M2$) comes along and offers two hundred dollars, in the absence of transaction costs it makes no difference for market efficiency if S decides to sell to $M1$ who resells to $M2$, or if S sells directly to $M2$ and voluntarily pays $M1$ one hundred dollars. Given the structure of the market for rough diamonds, however, if S sells to $M1$, it is unlikely that the stone will wind up in the hands of $M2$. Dealers keep their trading partners secret, particularly their sources of rough, since a dealer's ability to operate at a profit depends, in large part, on his network of contacts. After buying a stone that can be cut at a profit, most manufacturers do not want to incur the search cost of ensuring that the stone cannot be more profitably cut by another manufacturer, not do they want it known that they have a particular type of rough in their possession.

Given the remote possibility of resale, a rule that makes no allowance for the prospect of efficient breach in rough diamond sales may appear to induce too high a level of contractual performance. For a variety of reasons, however, the magnitude of the inefficiencies introduced by this rule is likely to be small, especially considering the increased importance of rough brokers in the market.

One major function of a broker is to conduct an effective search for the buyer willing to pay the highest price when the value of a stone is uncertain. If that uncertainty exists, the original owner has every incentive to hire the broker to search the market. If that has been done, then the first purchaser for use will rightly conclude that hiring a second broker has a very low rate of return, given the search already undertaken by the broker for the original owner. Similarly, if the original owner did not think it worthwhile to hire a broker for the original sale, unless circum-

stances have changed radically or the first purchaser has better information about the market, there is no reason to suppose that he will find it in his interest to either hire a broker or search the market himself.[61] In general, a rule requiring automatic performance will induce the optimal amount of search by sellers before the first sale is concluded. Thus, regardless of whether or not a broker was used in the original transaction, first purchasers will rarely find it advantageous to resell even if they search the market.

There is another reason that a rule of automatic performance does not introduce major inefficiency in the market. The cartel has the ability to fix the price of the rough that it sells. It also has a standard practice of announcing the magnitude of the price increase at each sight. Together, these two controls keep the difference between the prices that two manufacturers are willing to pay small relative to the aggregate benefit of avoiding the deadweight cost of dispute resolution. As an additional benefit, a high level of contractual performance in the sale of rough promotes efficient reliance decisions such as hiring skilled diamond cutters in advance to cut and polish the rough.[62] In aggregate, the magnitude of the inefficiencies introduced through a high level of contractual performance of executory promises to delivery rough stones is likely to be small, particularly since contracts for future delivery of a stone are uncommon and possession is typically transferred at the time of contracting. Thus, while it cannot be claimed that the rule of automatic performance in the sale of rough diamonds will always lead to the theoretically efficient outcome, the dynamics of the market suggest that the customary solution may well be the efficient solution when the imperfections brought on by positive transaction costs are taken into account.

Another problem associated with the use of extralegal contracts enforced through reputation bonds is the cost of renegotiation, which is

[61] In those situations where the first purchaser really does have superior connections for resale, he should enter the market as a middleman. This is observed: some of the largest manufacturers with the most extensive supply connections to sight holders, who often purchase rough stones in large parcels rather than individually, also run very active brokerage businesses.

[62] Diamond cutters are independent contractors and are often paid by the stone. Consequently, after contracting to purchase a piece of rough, a dealer will contract with a cutter. If he does not obtain the stone and does not have other work for the cutter to do, he will still have to pay the cutter. Furthermore, unlike many commercial contexts, at the time a diamond contract is made, the promisee typically is unable to estimate what is reliance expenditures will be; they will depend largely on the subsequent business opportunities that present themselves to the promisee. For example, if he subsequently promises to pay someone else and is unable to do so since he, himself, has not been paid, he will incur damage to his reputation and suffer a large loss.

likely to take place quite frequently since the "sanction [imposed in the event of breach] is much more likely substantially to undercompensate the promisee because implicit [extralegal] contract bonds often do not redound to the promisee's direct benefit."[63] Although the damage to the promisor's reputation does not directly redound to the benefit of the promisee, this problem has been largely overcome (at least with respect to transactions between club members) by the creation of the floor committee and the board of arbitrators, both of which have the authority to award damages.

G. The Substantive and Procedural Advantages of Arbitration over Adjudication

In the diamond industry, arbitration has important substantive and procedural advantages over adjudication. It enables parties to resolve disputes and enforce judgments quickly, inexpensively, and secretly, thereby containing damage to reputation and reducing the actual damage suffered by the promisee in event of breach.

Unlike courts, whose award of damages is limited by either expectation damages or a valid liquidated damages clause, the DDC bylaws allow arbitrators to award any measure of damages they think is appropriate, including punitive damages. They can also order one or both of the parties to pay a fine to a third-party beneficiary such as a charity. The authority to award punitive damages means that they can make the promisee whole, and the authority to order payment of a fine enables them to create a deterrent to breach contract. Since transactors know they may be forced to pay a penalty in the event of breach, their incentive to breach in the first place will be greatly reduced.

Although DDC arbitrators have industry expertise and sophisticated business judgment, they are not much better than courts at valuing lost profit or business opportunities forgone. Because arbitration hearings are held soon after the filing of the complaint, however, and because decisions are rendered and enforced shortly thereafter, the harm suffered by the promisee, while still difficult to quantify, is minimized. The inability of even expert dealers to accurately assess lost profit when a seller breaches a promise to deliver a stone may be the reason that possession is typically transferred at the time of contracting. Similarly, the difficulty of valuing lost business opportunities when a buyer fails to pay may account for the premium on speed: the sooner the promisee is paid, the

[63] Charny, *supra* note 58, at 5.

fewer transactions he will be required to forgo. The bourse's ability to resolve disputes promptly is considered so important that even if a dealer fails to appear for an arbitration, the hearing is held and he is bound by the panel's decision. The Floor Committee is also available during trading hours to resolve minor disputes as soon as they arise.

In disputes other than breach of a promise to pay money or deliver a stone, which are dealt with in the bylaws or according to well-established custom, arbitrators' verdicts may be more accurate and predictable than those of a court since arbitrators possess industry expertise and are permitted to consider information that would be excluded in court under the rules of evidence. If a diamond dispute were decided by a court, the application of industry custom would be highly unpredictable: unlike a DDC arbitrator, who can apply his own knowledge of industry custom, a judge would have to determine the content of customary norms from the conflicting testimony of expert witnesses. The uncertainty introduced by a judge's need to resolve conflicting testimony would greatly reduce the expected benefit to the promisee of having a legally enforceable contract.

Under the club's bylaws, the existence of a dispute and its resolution are kept secret so long as the arbitrators' judgment is paid promptly. Consequently, unlike filing a claim in court, initiating an arbitration does not affect the parties' ability to borrow or enter into implicit capital market transactions during the pendency of the dispute, which, in turn, minimizes the financial harm suffered by the promisee. The reputation damage suffered by the promisee is reduced by the practice of keeping disputes secret after a judgment is rendered since other transactors may view mere participation in an arbitration as a signal that a dealer was unwilling to renegotiate deals when unforeseen circumstances arose; they might demand additional protections or charge a higher price when dealing with him in the future.

The rapid enforcement of judgments is another advantage of DDC arbitration. Unlike a court, the DDC has the ability to bring unique pressures on the losing party to pay: it can put him out of business almost instantaneously by hanging his picture in the clubroom of every bourse in the world with a notice that he failed to pay his debt. Thus, the threat of publicity and the practice of keeping disputes secret as long as judgments are paid gives the defendant an incentive to promptly comply with the arbitrators' judgment. In addition, trade rules try to minimize the likelihood of a judgment-proof debtor in two ways: by making individual members as well as the corporations they trade for liable for arbitration judgments; and by providing for the expulsion of any member who files,

voluntarily or involuntarily, for personal or corporate bankruptcy in court instead of going through the club's own bankruptcy procedure, which requires the debtor to make 100 percent restitution to his diamond industry creditors.

Although keeping this type of information about dealer behavior from a market that works largely on reputation may slightly impair the efficient operation of the market for reputation information, in the context of the diamond industry's institutional structure, there are sound reasons for this practice. Requiring arbitration judgments to be made public without introducing additional changes in the system might result in the dissemination of information that would be difficult for the market to value accurately. If only the amount of the judgment were announced, a dealer who was ordered to pay a large judgment because there had been an honest misunderstanding in a large transaction would suffer more reputational damage than a dealer who had to pay a smaller judgment because of deliberate breach or theft. Consequently, the facts of the case would have to be released to accurately convey the relevant information to the market if judgments were made public. Arbitrators would have to make findings of fact and issue written opinions, which would lead to a demand for procedural protections such as rules of evidence and more extensive discovery. In time, the flexibility and informality of the system, essential to the rapid resolution of disputes, would begin to disintegrate. Furthermore, it may be that the information most important to the reputation market is not that a dealer has been involved in a dispute or even that he has breached a contract, but rather that he has been prepared to either settle disputes or abide by the judgments of the arbitral tribunal when a third-party adjudication was necessary.

In complex cases not covered by the trade rules or industry custom, diamond industry arbitration suffers from the same weakness as most commercial arbitration: unpredictability. The lack of written decisions and a tradition of stare decisis makes it difficult for market participants to make rational breach decisions and to determine in advance the type of sanctioned behavior. In order to increase predictability, many bourses in the world federation have relaxed the norm of complete secrecy. Arbitrators publish written announcements of the principles used to decide novel cases while keeping the parties and other identifying facts secret. The WFDB recently proposed compiling a computer data base of these statements of principle to promote worldwide uniformity of arbitrated judgments and to prevent "forum shopping." They also proposed additional uniform training programs for all arbitrators. Younger WFDB officials fear that if such changes are not introduced the system will be

perceived as arbitrary and unjust, and its legitimacy may decline. Recently,[64] there has been increasing pressure on the New York bourse to relax the secrecy norm and to permit arbitrators to issue policy statements in novel or complex cases—a change that would enable the industry to capture the benefits of arbitration (secrecy, informality, and speed) and litigation (the creation of precedent and stare decisis).

H. The Aggregate Efficiency of the System

The importance of international transactions suggests that concluding transactions in accordance with a nearly uniform system of private law has additional efficiency benefits. If a dealer is a member of any one bourse in the world federation, he is automatically admitted to the trading floor of all of member bourses. Most diamond dealers frequently transact in foreign bourses. It would be wasteful for dealers to have to learn the trade rules of different bourses and be concerned with the technicalities of concluding legally enforceble agreements in different countries, particularly when many of these countries do not have well functioning judiciaries. The world federation maintains a board of arbitrators that has the authority to settle disputes between bourses or to hear cases between private litigants from different bourses when there is a colorable question as to which party's bourse should hear the case. Resolving disputes through private international arbitration also avoids complex questions of international jurisdiction.

[64] A few years ago a case arose that revived the debate over the secrecy of judgments in the New York bourse. The Yehuda treatment is a way of altering a stone such that its flaws become invisible to the human eye unaided by special technology. The firm that developed this process and actually treats the stones requires those they deal with to sign an agreement requiring disclosure of the stone's treatment to any potential buyers. Soon after the treatment was introduced, but before it was widely known, a dealer sold a treated stone without disclosing the treatment. The buyer subsequently discovered the treatment and filed a claim against the seller. The seller defended on the grounds that he did not know or have reason to believe that the stone had been so treated. The board of arbitrators ordered recision of the deal and imposed a very small fine on the seller. One arbitrator wanted to write an opinion explaining that the only reason the judgment was so small was that the treatment was new and a dealer in exercise of ordinary care would not have been expected to ask whether or not the stone had undergone this treatment. By the time the arbitration was concluded, however, the treatment had become so well known that a similar defense of ordinary care would not prevail in the future and the arbitrators intended to impose extremely heavy fines in subsequent cases. Some members of the DDC board of arbitrators are concerned that the lack of published opinions explaining the basis of decisions gives dealers the wrong signals about what type of behavior is sanctioned. Although cases are officially kept secret, the industry is "like a bunch of old ladies," and in new and unusual cases the result can rapidly become known.

I. The Importance of Reputation Bonds in the Market as a Whole

Diamond dealers consistently maintain that transactions between two club members, between two nonmembers, and between a member and a nonmember are conducted in exactly the same way. If the availability of the DDC's arbitration system and enforcement mechanisms were central to parties' ex ante decision making, the terms of the transaction (either substantive or price) should be different when at least one party is a nonmember. For example, a member seller who would sell a stone to another member on thirty-day terms would be expected to charge a nonmember a higher price (or perhaps demand cash on the spot) to compensate for the risk of nonpayment and the unavailability of arbitration. Dealers insist, however, that no such differences exist and that they decide who to deal with purely on the basis of the other party's reputation.

If reputation bonds are well functioning, this behavior is not surprising. In a transaction between a member and a nonmember, the nonmember has an incentive to keep the bargain if he wants to be admitted to the bourse in the future. The economic benefits of bourse membership make it actively sought after by most market participants. A member can not only spread the word about the nonmember's wrongdoing, but he can also object to his being accepted for club membership. In transactions between two nonmembers, both parties have reason to worry about their reputations. In order to obtain a steady supply of rough to run an efficient manufacturing business, a nonmember must have a reputation of being scrupulously trustworthy. Nonmembers know that their potential future trading partners will inquire more deeply into their reputation before transacting with them since they do not have the club's stamp of approval.

If dealers really did rely on arbitration to resolve most disputes, one would expect that if it were not available more disputes would go to court. This is not observed; litigation between two nonmembers is also infrequent. Similarly, if reputation bonds were not strong enough to enforce arbitration judgments, one would expect to see frequent recourse to the courts for judicial confirmation of arbitrated judgments. This is also not observed. Thus, it appears that the dispute resolution institutions in the diamond industry can fairly be called extralegal: it is primarily the fear of damage to reputation that maintains discipline in the diamond trade, not the bourse's board of arbitrators or the procedural right to appeal arbitrated decisions in court.

The relative importance of reputation and arbitration may now be shifting, for dealers differ among themselves on the importance of arbitration's availability. Most claim that it is unimportant, but there are recent

signs to the contrary. In the early 1980s, one reason dealers gave for leaving the newly formed Los Angeles club was that it did not provide arbitration. Furthermore, the president of the World Federation of Diamond Bourses is concerned that, as trust breaks down and dealers become increasingly focused on their "rights," arbitration will come to have a more important function. He believes that the increasing importance of arbitration and third-party dispute resolution requires more qualified arbitrators and greater uniformity of decisions and is concerned that, unless the bourses meet the challenge of providing a quick and predictable way of resolving disputes, the diamond industry's independence from the legal system will slowly disintegrate.

V. Arbitration and Settlement

With respect to simple disputes dealt within the bylaws or those dealt with according to well-established custom, the decision whether to settle or go to arbitration will depend on the usual parameters. The expected value of the arbitration to the plaintiff will be the probability of success on the merits, multiplied by the projected recovery, less the cost of legal representation if represented by counsel, less (depending on the arbitrator's whim) the cost of arbitration if he is made to bear it. The bylaws provide that the plaintiff must pay the arbitration fee in the first instance but give the arbitrators the discretion to refund the fee or order the dependant to pay it. Although this fee-shifting term is a wild card, it is bounded by the actual cost of arbitration, which is quite low relative to the amounts at stake in the arbitration.

Conversely, the expected cost of the arbitration to the defendant is the probability of losing multiplied by the damage award, plus legal fees if represented by counsel, and, perhaps, the cost of the arbitration if the arbitrators, in their discretion, order him to pay it. Models of suit and settlement[65] suggest that the closer the plaintiff and defendant's estimates of the expected outcome of the litigation, the more likely they are to settle. Consequently, to the extent that the required prearbitration conciliation proceedings shed light on the strengths and weaknesses of the parties' arguments, they would be expected to lead to a high rate of settlement. This is, in fact, observed: 80–85 percent of the disputes submitted to arbitration are settled during the proceeding's mandatory conciliation phase.

In more complex cases—such as labor disputes, trademark infringe-

[65] See Steven Shavell, Suit, Settlement, and Trial: A Theoretical Analysis under Alternative Methods for the Allocation of Legal Costs, 11 J. Legal Stud. 55 (1982).

ments, partnership disagreements, and the use of new techniques to make flaws in stones invisible to the human eye—a party cannot simply be ordered to pay the money owed or to deliver or return the stone in question. In these cases, the arbitration panel either hears the case, or, if it falls into one of the four categories enumerated in the bylaws,[66] the parties are left free to seek a resolution of their dispute in court. When arbitrators opt to decide complex or novel cases, however, it is difficult for the parties to predict the rule of decision and/or the damage measure that arbitrators will apply. Since arbitrators neither make findings of fact nor render written opinions announcing their decisions, past decisions are a poor predictor of future outcomes. As a consequence of both parties' inability to predict how the arbitrators will decide complex cases, in situations where the parties do not differ greatly in their degree of risk aversion and have similar estimates of the degree of uncertainty in the arbitrators' decisions, they also will have an incentive to settle, just as they did when they had near-perfect information about the rule of decision and the damage measures that the arbitrators would employ were certain.

VI. The Effect of Legal Intervention into the Extralegal Contractual Regime

In general, diamond dealers prefer to conclude agreements using extralegal contracts. Certain types of agreements made in the course of diamond transactions, however, are routinely subject to interpretation by the courts since they often involve the rights of third parties. Consequently, these agreements often take the form of legally enforceable contracts.

One example is when a bank or an insurance company is a direct party to an agreement. Unlike individual buyers and sellers, banks and insurance companies do not have an interest in maintaining the secrecy norm. These institutional actors often have significant bargaining power, particularly the banks, since in most countries a relatively small number of banks provide most of the industry's financing.[67] Consequently, banks

[66] See text around note 25 *infra*.

[67] One reason a relatively small number of banks are involved in the diamond industry is that evaluating the worth of a stone (often used as inventory collateral) in the absence of an objective and readily ascertainable market price requires an expertise in gemstones that bankers seldom have. Consequently, many loan decisions are really made on the basis of the bank's perception of the dealer's reputation in the marketplace. As an officer of the Merchants Bank of New York (located in the middle of Forty-seventh Street) explained, "[i]n terms of extending credit a bank has to look at the 3 C's—Capital, Culpability, and Character. At our bank, we think that character is the most important C." Merchants Bank

are often able to obtain the benefit of having a legally enforceable contract, such as a standard loan agreement, as well as the implicit collateral of a reputation bond. A second example is found in transactions such as consignments, where banks or insurance companies are not directly involved, but where their rights may be affected later and the legal process invoked to resolve disputes.

Consignment agreements used to be concluded orally. Under the trade rules for consignment, title to the goods remained in the owner, and he was entitled to get them back if they were not sold on his behalf. The courts have been reluctant to credit arguments based on custom and usage, however, and generally have refused to recognize the existence of the extralegal agreement to return the goods, finding them to be the property of the consignee. As a consequence, when a consignee goes bankrupt, courts do not permit the consignor to recover his diamonds: "diamonds delivered on memo to a broker or dealer usually cannot be recouped from a trustee in bankruptcy, an assignee for the benefit of creditors or even from a bank from whom your consignee has borrowed money and given his bank the normal and usual security interest in his inventory and accounts."[68] Consequently, when a dealer gives goods on consignment, a formal consignment memorandum that satisfies the requirements of the Uniform Commercial Code is now sometimes drawn up to ensure that the dealer's title to the goods will be recognized by the legal system.[69] Dealers explain that the documents are designed to serve two distinct purposes. Between the dealers, their function is similar to that of the bill of sale, weight slip, or cachet wrapper—they are intended to help the dealers privately settle any disputes that may arise by clarifying the terms of the original agreement. These agreements are not drawn up in the form of legally enforceable contracts because the dealers think the consignee will abscond with the goods. The same risk of loss would be present in any sale for future payment (especially since consignment

---

Moved and Grew with Industry, N.Y. Diamonds, December 1988, at 38. Thus, although defaulting on a loan would hurt any businessman's credit rating, the damage to a diamond dealer is more severe since there are only a few industry lenders and banks must rely to a greater extent on the dealers' reputation in valuing his assets.

[68] S. Herman Klarsfeld, Legal Gems, N.Y. Diamonds, May 1988, at 40.

[69] As the club's legal counsel recently advised dealers "the Uniform Commercial Code (UCC) will give you protection if you adequately describe your diamonds and file a UCC-1 Financing Statement with the Secretary of State in Albany and with the register of the county in which the consignee has an office. . . . This will give you a legal leg to stand on if you unfortunately have to seek the return of your merchandise from a bank or a trustee in bankruptcy." *Id.* at 63. However, due to the transactions costs of drafting and filing the financing statement, they are used only in the largest transactions.

agreements and sales are often made between the same people), a situation in which dealers clearly prefer extralegal agreements. Legally enforceable contracts are sometimes used in consignments because these transactions are often interpreted in the course of legal proceedings, and without them courts tend to interpret the meaning of an intraindustry consignment agreement in ways that are strongly at odds with industry custom and the intent of the original contracting parties.

Throughout its history, the diamond business has been largely self-regulating, operating outside the law of the state. Over the past thirty-five years, the private dispute-resolution mechanisms in the world's diamond bourses, combined with widespread adherence to the secrecy norm, have succeeded in maintaining a largely extralegal contractual regime where transactions are concluded on the basis of the dealers' reputations and the incidence of breach is low.

Over the past decade, however, a subtle change has been taking place—the legal system has begun to interfere with the substantive rules used to decide arbitrated cases as well as the ways in which these decisions are enforced. Under the DDC bylaws, the Board of Arbitrators can suspend or expel a member if he does not comply with a judgment. Ever since Martin Rapaport[70] decided to break the secrecy norm, however, by initiating the first suit against the club for any reason other than disagreement with an arbitration decision, there has been a profound change in the way the club decides cases and enforces judgments. The Rapaport controversy has made the club much more reluctant to expel members—it is concerned not only about the expelled member bringing suit, but it also fears that too many expulsions will revive the Federal Trade Commission's interest in its activities. At present, a member is not expelled until the Board of Arbitrators first obtains a court order affirming its decision. Effective sanctions may still remain, however, since the member's picture, along with a description of the judgment that he refused to pay will still be hung in the club room and on the trading floor of every bourse in the world federation.

Although the DDC bylaws have always given the litigants the right to be represented by a lawyer, prior to the Rapaport case it was uncommon. Today, legal representation is the norm. The arbitrators feel that the presence of lawyers has, in some measure, altered the rules of decision they apply. The lawyers alert them to relevant parts of New York law, and, while this law still does not supply the rule of decision, the arbitrators are more conscious of the law and are increasingly reluctant to drastically depart from it, except in instances where the decisions are deeply

[70] See note 50 *supra*.

rooted in custom or do not involve creating a new rule. Although the Board of Arbitrators has traditionally declined jurisdiction in cases involving complex statutory rights or claims that are intertwined with pending litigation, in recent years, this has become a more common practice. The older arbitrators fear that legal interference in the diamond trade will one day destroy the traditional way of doing business.

VII. Conclusion

This article has been largely devoted to offering explanations of why the diamond industry has long relied on the extralegal enforcement of its business norms. By a variety of reputational bonds, customary business practices, and arbitration proceedings, the diamond industry has developed a set of rules and institutions that its participants find clearly superior to the legal system. The industry, as it has been traditionally organized, is able to make and, more important, enforce its own rules. The market is organized to promote the low cost and rapid intraindustry dissemination of information about reputation, which enables it to use reputation bonds to create intraindustry norms that function as a deterrent to breach of contract and a private sanctioning system whose judgments can almost always be enforced completely outside the legal system.

The customs and institutions in the diamond industry emerged for reasons wholly unrelated to shortcomings in the legal system; yet, even as the force of the old enforcement mechanisms of religion and secondary social bonds began to disintegrate, a network of trading clubs, designed to promote the dissemination of information about reputation and socialization among members, emerged to fill the gap. That generations of diamond dealers have clung to nearly identical intraindustry norms in countries with a wide variety of legal rules and institutions suggests that the traditional rules and institutions are likely to be efficient from the perspective of market insiders. In the United States, the traditional rules and institutions endured over time and demonstrated their superiority to the established legal regime.

In the diamond industry, "trust" and "reputation" have an actual market value. As an elderly Israeli diamond dealer explained, "when I first entered the business, the conception was that truth and trust were simply *the* way to do business, and nobody decent would consider doing it differently. Although many transactions are still consummated on the basis of trust and truthfulness, this is done because these qualities are viewed as good for business, a way to make a profit."[71]

[71] Interview with author, summer 1989.

This page intentionally left blank

Acknowledgments

Cohen, Morris R. "The Basis of Contract." *Harvard Law Review* 46 (1933): 553–92. Copyright 1933 by the Harvard Law Review Association.

Kessler, Friedrich. "Contracts of Adhesion — Some Thoughts About Freedom of Contract." *Columbia Law Review* 43 (1943): 629–42.

Dawson, John P. "Economic Duress — An Essay in Perspective." *Michigan Law Review* 45 (1947): 253–90.

Letwin, William L. "The English Common Law Concerning Monopolies." *University of Chicago Law Review* 21 (1954): 355–85. Reprinted with the permission of the University of Chicago Law School.

Leff, Arthur Allen. "Unconscionability and the Code — The Emperor's New Clause." *University of Pennsylvania Law Review* 115 (1967): 485–559. Reprinted with the permission of the *University of Pennsylvania Law Review* and William S. Hein and Company, Inc. Copyright 1967 University of Pennsylvania Law Review. All rights reserved.

Blades, Lawrence E. "Employment at Will vs. Individual Freedom: On Limiting the Abusive Exercise of Employer Power." *Columbia Law Review* 67 (1967): 1404–35. Reprinted with permission.

Epstein, Richard A. "In Defense of the Contract at Will." *University of Chicago Law Review* 51 (1984): 947-82. Reprinted with the permission of the University of Chicago Law School.

Radin, Margaret Jane. "Market-Inalienability." *Harvard Law Review* 86 (1987): 1849–1937. Copyright 1987 by the Harvard Law Review Association.

Bernstein, Lisa. "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry." *Journal of Legal Studies* 21 (1992): 115–57. Reprinted with the permission of the University of Chicago Law School. Copyright by The University of Chicago. All rights reserved.