Norms in Surprising Places: The Case of Statutory Interpretation*

Cass R. Sunstein

Philosophers, political scientists, and lawyers have devoted considerable attention to the role of norms in the two areas for which Anglo-American legal systems are most frequently celebrated: the common law and constitutional law. Thus, for example, judicial elaboration of common law principles of contract, tort, and property has been described (alternately) as an effort to promote economic efficiency or to protect private autonomy; and background norms of various sorts have been developed from these premises. So too, the role of norms in the choice among different strategies for interpreting the open-ended phrases of the Constitution is reasonably well understood.

There is a conspicuous gap, however, in the treatment of statutory interpretation. The gap is striking in light of the fact that the interpretation of statutes has become by far the most frequent and important task for judges (and administrators) in industrialized democracies. Although interpretation of regulatory statutes frequently determines the content and scope of programs affecting (for example) pollution, energy, nuclear power, endangered species, workers, and victims of discrimination, it is in this setting that courts are said to have little or no room to move. On the usual view, their role is to act as agents of the legislature, implementing its commands. Indeed, it is here that courts are said to act, at least most of the time, without "interpretation" or "norms" at all. In these circumstances, the notion that norms play a major role in statutory construction might seem iconoclastic or even bizarre.

In this essay, I want to argue that extratextual norms—understood as principles about constitutional government, institutional arrangements, basic fairness, and regulatory failure—do in fact play a crucial role in the interpretation of statutes. Indeed, descriptive and prescriptive work on this topic is impossible without an understanding of the role of norms. My argument comes in two basic parts. In the first part, I reject the view

* Many of the arguments in this essay are given in more detail in a forthcoming book, *After the Rights Revolution: Reconceiving the Regulatory State* (Cambridge, Mass.: Harvard University Press, 1990), and also in "Interpreting Statutes in the Regulatory State," *Harvard Law Review,* vol. 103 (December 1989).

Ethics 100 (July 1990): 803–820

that courts can be "agents" of the legislature and suggest instead that extratextual norms are a necessary part of statutory interpretation. In the second part, which is more speculative, I describe possible foundations for extratextual norms and outline their role in statutory construction. Above all, I urge that it is desirable, not to limit the role of such norms, but to identify them, to evaluate them, and to develop and apply norms that are well suited to, and improve the performance of, modern regulatory systems.

COURTS AS AGENTS

Textualism

It is sometimes suggested that the statutory language is the only legitimate basis for interpretation. The principal feature of the textualist creed is the belief that courts have no license to rely on "outside sources." A particular target here is the use of norms that cannot be tied to statutory text itself. The language of the statute—and nothing else—is the source of judicial power and the object of judicial concern.

Textualism appears to be enjoying a renaissance in a number of recent cases, and perhaps in the academy as well. Whatever its precise form, however, the textualist approach is inadequate. The central problem is that words are not self-defining; their meaning depends on both culture and context. There is no such thing as a preinterpretive text, and words have no meaning before or without interpretation. To say this is not to say that words used in statutes can mean anything at all. (Confusion of the two quite separate points has been frequent in recent work on literary interpretation and on linguistics.) But it is to say that statutory terms are indeterminate standing "by themselves," and, more important, that they never stand by themselves. For this reason, background norms of interpretation, orienting the judicial reader to the text but not "in" it, are an indispensable part of interpretation.

Usually the background norms are so widely shared—for example, that Congress is speaking in English, that Congress is not joking or attempting to mislead, or that courts should not decide cases simply according to their predilections—that they are invisible. But in easy as well as hard cases, courts must resort to such norms if interpretation is to proceed. For example, the question whether the enactment of federal environmental statutes preempts all of state tort law is an easy one—it does not—not because of the statutory text "itself" but because of universally shared understandings about the limited preemptive effect of federal enactments.

So far, perhaps, those who accept textualist approaches might be in agreement. Perhaps they would agree that the meaning of words is largely a function of background norms about how texts should be approached. Indeed, we may define an easy case as one in which there is a consensus about the governing norms. But the problem for textualism goes deeper.

In many hard cases, the source of the difficulty is that the particular kind of background norm and the nature of its application will be highly controversial.

Assume, for example, that the question is whether a statute creating a regulatory agency to bar racial discrimination in employment implicitly authorizes private suits by victims of discrimination against employers who have engaged in discrimination. If the statute and its history are silent on the question, the case cannot be decided without introducing background norms, which are necessary to tell interpreters what such silence means. Thus it should be unsurprising that judges skeptical about implied causes of action have relied, not on statutory "text," but on a background assumption, traceable to a particular understanding of separation of powers, that only the legislature may create such rights, and that statutory silence should be assumed not to do so.

These considerations account for the pervasive difficulties with textualist approaches to statutory construction: ambiguity, delegation or gaps, overinclusiveness, underinclusiveness, and changed circumstances. The most familiar problem is that statutory language is sometimes ambiguous. It is not clear, for example, whether the term "feasible" contemplates cost-benefit analysis, whether a statutory reference to "wages" is meant to include employer contributions to a pension program, whether regulatory power over the broadcasting industry includes cable television, or whether a prohibition of "discrimination" bars voluntary race-conscious measures designed to counteract the effects of past and present discrimination against blacks.[1] In all of these cases, moreover, it is uncertain whether the language should be taken to refer to the original meaning of those words for the enacting legislature (assuming that idea can be made intelligible in light of the problems of aggregating the viéws of numerous actors) or should instead take account of contemporary understandings of what the words mean. By itself, textualism cannot answer these questions.

Statutes might also represent a delegation of lawmaking power to courts, or contain gaps that interpreters must fill. For example, the basic terms of the most important American antitrust law—forbidding "conspiracies in restraint of trade"—are uninformative about the sorts of norms that might be brought to bear in deciding precisely which conspiracies are unlawful. So too, the statutory prohibition on "discrimination" in civil rights law is silent on burdens of pleading and production, the appropriate statute of limitations, and a range of additional implementing rules. In such cases, the problem is that the text requires interpreters to look to sources outside of its four concerns.

Sometimes, language that would in many contexts be entirely unambiguous should not be the only basis for interpretation. The literal

1. See, e.g., Industrial Union Dep't, AFL-CIO v. American Petroleum Inst., 448 U.S. 607 (1980) (meaning of "feasible"); Steelworkers v. Weber, 443 U.S. 193 (1979) (discrimination).

language of a statute—the meaning of its terms in ordinary settings—will suggest an outcome that would make little or no sense. Suppose, for example, that a state law says that no vehicles are permitted in public parks and that a city proposes to build in a park a monument consisting of tanks used in World War II.[2] Here what appears to be the unambiguous language must yield, for the statute could not reasonably be taken to forbid such a monument, which causes none of the harms the statute could be thought to prevent.

The example is not fanciful. The Supreme Court recently said that a statute exempting state and local public housing agencies "from all taxation . . . imposed by the United States" should not be interpreted to include an exemption from federal estate tax.[3] The Court held that the exemption did not mean what it appeared to say in light of the contemporaneous understanding that an excise tax was not ordinarily comprehended within the category of "taxation." Or suppose that the legislature has said that an employer may discharge an employee "for any reason." Is the employer thereby authorized to fire workers who have refused to commit crimes on his behalf? Might a man who has murdered his wife inherit from her under a statute allowing spouses to inherit "under all circumstances"? These are examples of what might be described as the overinclusiveness of textualism: the possibility that statutory language, read without regard to context, will reach situations that it ought not to cover.

Courts encounter the problem frequently.[4] A passage from Wittgenstein indicates the basic difficulty: "Someone says to me: 'Shew the children a game.' I teach them gaming with dice, and the other says 'I didn't mean that sort of game.' Must the exclusion of the game with dice have come before his mind when he gave me the order?"[5] The example shows that sometimes the best interpretation of a textual command is one that runs counter to its apparent literal meaning—even in cases in which the author did not have in mind the particular case at issue or make a judgment about how that case should be resolved.

Although it arises less frequently, there is also a possibility that textualism will be underinclusive. This problem occurs when the text, read outside of context, does not regulate conduct that the statute, most

2. The example comes from H. L. A. Hart, *The Concept of Law* (Cambridge: Cambridge University Press, 1961), pp. 121–32.

3. United States v. Wells Fargo Bank, 108 S.Ct. 1179 (1988).

4. See California Federal Savings & Loan v. Guerra, 479 U.S. 272 (1987); O'Connor v. United States, 479 U.S. 27 (1986); Offshore Logistics, Inc. v. Tallentire, 477 U.S. 207, 220–21 (1986); Holy Trinity Church v. US, 143 U.S. 457, 459 (1892); Steelworkers v. Weber, 443 U.S. 193, 201 (1979); Kelly v. Robinson, 479 U.S. 36, 43–44 (1986); Perry v. Strawbridge, 209 Mo. 621 (murderer may not inherit as a result of death caused by his crime, notwithstanding clear statutory language).

5. L. Wittgenstein, *Philosophical Investigations,* trans. G. Anscombe (Oxford: Blackwell, 1972), p. 33.

plausibly read, should be understood to control. A particular difficulty here is that a statute may be "evaded" by private ingenuity. The literal language of the statute does not cover the situation, but because the private conduct causes all of the harms that the statute prevents, courts sometimes hold a statute applicable notwithstanding its precise terms.[6] Consider Justice Holmes's suggestion that courts are "apt to err by sticking too close to the words of a law where those words import a policy that goes beyond them."[7]

The discussion thus far has assumed that circumstances have not changed significantly since the statute was enacted. But the difficulties of textualism become all the more severe when time has affected the assumptions under which the statute was originally written. Consider the Delaney Clause, which forbids the use as food additives of substances that "induce" cancer. The Delaney Clause was enacted at a time when carcinogenic substances were difficult to detect, and all detectable carcinogens were extremely dangerous. Currently, carcinogens are detectable at exceptionally low levels, and many of them pose trivial risks. In such circumstances, might the word "induce" allow the government to exempt from regulation carcinogens that pose trivial risks? Mechanical application of the text in new circumstances is likely to produce interpretive blunders.

All this suggests two principal points. First, and most fundamentally, there is no such thing as an acontextual "text" that can be used as the exclusive guide to interpretation. In easy cases, background norms—on which there is wide or universal consensus—are part of the process of ascertaining statutory meaning. Precisely because the norms are so widely shared, they appear invisible and are not an object of controversy. In other cases, usually treated as hard ones, courts must resort to a contestable background norm in order to resolve an interpretive dispute. It is in these settings, and in this sense, that textualism is inevitably incomplete.

Second, it is by no means obvious that courts should always rely on the text or on the "plain meaning" of words even in cases in which such reliance is possible and leads to determine results. An interpretive strategy that relies exclusively on the ordinary, dictionary definition of words is precisely that—an interpretive strategy that reflects a choice among competing possibilities—and in many cases it will produce irrationality, absurdity, or an inferior system of law. The invocation of background norms of various sorts promises to be a salutary corrective.

The Inevitable Failure of the Agency View of the Judicial Role

Thus far we have seen that the most prominent species of the agency view of the judicial role appears to fail. It is time to see what conclusions emerge from that failure. On what remains the most well known approach to the matter, the process of statutory construction involves ascertaining

6. See, e.g., Helvering v. Gregory, 69 F. 2d 804 (2d Cir. 1934).
7. Olmstead v. United States, 277 U.S. 438, 469 (1928).

the meaning of the legislative command in the particular case. This approach sees the courts as agents or servants of the legislature.

On the agency view, courts should say what the statute means, and in that process language, history, and structure are relevant; and perhaps those who accept the agency view, rejecting textualism, would concede the need for interpretive principles to help orient reader to text. But on this view, background norms of other sorts—taken up below—are impermissible. The agency view amounts to a formalist approach to statutory construction—formalist because it does not permit courts to undertake value-laden inquiries into statutory function and failure as part of the process of interpretation.

The agency view is usually defended by a claim of legitimacy. In a democratic system, one with separated powers, it is said to be impermissible for courts to invoke norms that cannot be traced to an authoritative textual instrument. The claim of illegitimacy is buttressed by (or reducible to) a range of prudential considerations: the use of background norms will tend to increase judicial discretion, decrease legislative attentiveness, produce uncertainty, and risk usurpation by judges of powers accorded to legislative and executive actors. In this sense, the debate over formalism in statutory construction replicates a similar debate in constitutional law. As a particularly crisp expression of the agency view of the judicial role, consider the words of Justice Holmes, who, after describing a regulatory statute as "foolish," added: "If my fellow citizens want to go to Hell I will help them. It's my job."[8]

The agency view contains an important truth. It would be improper for courts to "interpret" statutes by construing them to mean whatever the judges think would be best. No one could defend an approach to statutory construction that would license judges to say that a statute means whatever a good statute would say. It follows from this understanding that where there is neither interpretive doubt nor constitutional objection, it is the judgment of the electorally accountable branch, and not that of the judges, that will prevail.

Notwithstanding these considerations, the formalist position, and the agency view of statutory construction, are subject to two decisive objections. First, the formalist conception of "meaning" is far too crude. As we have seen, language is not self-interpreting. Courts cannot "simply" be agents. The instructions of the principal are unintelligible without background norms that interpreters alone can supply. It is important to emphasize that this is a conceptual or logical claim, not a proposition about the appropriate distribution of powers as among administrators, courts, and legislatures.

Suppose this much were acknowledged, and it were suggested that courts should rely on the agency view "as much as possible." This is a

8. *Holmes-Laski Letters,* ed. M. Howe (Cambridge, Mass.: Harvard University Press, 1953), vol. 1, p. 249.

highly ambiguous concept, but perhaps it can be taken to argue in favor of adherence to legislative meaning to the extent that it is fairly ascertainable for judges equipped with conventional or uncontested background norms. The second problem for the agency view is that its claim of illegitimacy relies on question begging and probably indefensible premises. That claim must be defended by a set of institutional and substantive arguments, and those arguments are hard to supply. The claim must be that a formalist approach to construction—again, to the extent that it is possible—will lead on balance to the best or most sensible system of law. But it is far more likely that the introduction of other considerations will accomplish that goal. Assume, for example, that statutory language, while ambiguous, is most naturally read to lead in a direction that is, by general consensus, absurd and could not plausibly have been intended; that a statute might be understood as giving open-ended authority to a bureaucracy; that the most obvious reading of a statute provides bizarre or unjust results in light of changes in the fifty years since its enactment; that an appropriations measure appears to intrude on a carefully elaborated (and sensible) compromise in the environmental field. In all of these settings, invocation of background norms will lead to better outcomes. In such cases, courts that act as something other than mere agents will produce a more sensible system of law.

THE ROLE OF INTERPRETIVE PRINCIPLES

The Unlikely Promise of "Canons"—and an Analogy

How, in light of these considerations, ought the problem of statutory construction to be understood? Some help in this regard might be found in an unlikely place. In interpreting statutes, courts have often relied on "canons" of construction. I will use the term "canons" to refer to all background norms that are used in statutory construction.

Any defense of the canons must come to terms with one of the most important (and characteristic) contributions of the legal realist movement, which was to discredit the canons. The most important example is Karl Llewellyn's celebrated effort to demonstrate that for each canon there is an equal canon pointing in the opposite direction.[9] In Llewellyn's view, an inspection of the decisions revealed that the canons did not help to decide cases; instead they operated as mechanical, after-the-fact recitations disguising the reasons for decision. This view of the canons has deeply penetrated modern legal culture. Almost no one has had a favorable word to say about the canons in many years.

But the canons of construction continue to be a prominent feature in the federal and state courts. The use of guides to construction—in the form of background understandings and principles requiring a "clear statement" from Congress in various settings—can be found in every

9. See K. Llewellyn, "Remarks on the Theory of Appellate Decisions and the Rules or Canons about How Statutes Are to Be Construed," *Vanderbilt Law Review* 3 (1950): 395.

area of modern law. And there is no sign that canons are decreasing in importance—partly, as we have seen, because of their inevitability and partly because of their plausibility on substantive grounds. The deviation between judicial practice and academic consensus is something of a puzzle.

An analogy may be helpful here. The law of contracts is pervaded by—indeed, it consists largely of—a set of principles filling in contractual gaps when the parties have been silent, or when the meaning of their words is unclear. The use of implied terms is a familiar part of the law of contract; and it would be most peculiar to say that they are an illegitimate incursion into the usual process of "interpreting" the parties' intent. Without implied terms of some sort, contracts simply would not be susceptible to construction.

To a large degree, interpretive norms serve the same function in public law. They too help judges to construe both statements and silences; they too should not be seen as the intrusion of controversial judgments into "ordinary" interpretation. But there are differences as well as similarities. In the law of contract, it is often said that implied terms should attempt to "mimic the market" by doing what the parties would do if they had made provisions on the subject. In the law of statutory construction, by contrast, the notion of "mimicking the market" is far from a complete guide. Sometimes it will be extremely unclear how Congress would have resolved the question; sometimes the resolution of the enacting Congress would have been different from that of the current Congress, and it is not clear whose resolution should control; sometimes courts properly call into play principles—many of them constitutionally inspired—that push statutes in directions that diverge from the conclusion that Congress would have reached if it had resolved the matter.

Despite these differences, the analogy to contract law reveals that it is impossible to dispense with background norms in interpreting terms. In the context of statutory construction, however, such norms will often serve functions other than that of applying the judgment that Congress would have reached if it had addressed the question.

The Functions of Interpretive Principles: A Catalog

Part of the problem is coming to terms with the canons, or with background norms of interpretation, stems from the fact that they serve a variety of different goals. Those goals fall in four basic categories.

1. *Orientation to meaning in particular cases.*—Interpretive norms may be designed to orient judicial readers to the text in order to help them to ascertain its meaning in the particular case. This is the most uncontroversial function of interpretive principles, though not all of them sensibly promote this goal. Consider, for example, the idea that the language of a particular provision will be taken in the context of the statute as a whole and will not be interpreted so as to do violence to statutory structure. The understanding that even apparently unambiguous statutory terms might

have a counterintuitive meaning if other provisions of the statute so indicate is a sensible way of understanding statutory meaning in the particular case.

2. *Interpretive instructions.*—Interpretive norms may serve as guides to what might be called the interpretive instructions of the legislature. Such principles are designed to capture an actual or hypothetical legislative judgment about how statutes should be construed. They do not serve as guidance to meaning in the particular case but, instead, attempt to incorporate explicit or implicit congressional commands with respect to the interpretive process.

The easiest cases here involve explicit legislative instructions about interpretation. Thus the first sections of the United States Code set out a series of guidelines for courts to follow in interpreting statutes.[10] When Congress has been explicit, there can be no objection to judicial use of the relevant instructions. Some interpretive principles, however, are a product of an understanding of implicit rather than explicit legislative interpretive instructions. Consider the idea that statutes will be interpreted so as to be constitutional rather than unconstitutional, or that judicial review of administrative behavior will be assumed in the absence of a clear congressional instruction to the contrary.

3. *Improving lawmaking.*—A third function of interpretive norms is to promote better lawmaking. Such principles might, for example, minimize judicial or administrative discretion or push legislative processes in desirable directions. Principles of this sort attempt neither to orient the reader to the text in the particular case nor to capture congressional instructions with respect to interpretation. The effort is instead to improve lawmaking processes and the deliberation and accountability that are supposed to accompany them. In this respect, some interpretive principles fulfill goals associated with the separation of powers and with plausible assessments of comparative institutional competence. They are designed above all to channel certain decisions through certain institutions or to improve the operation of those institutions. Examples include the idea that appropriations statutes, usually done hastily, will not be understood to amend substantive statutes; that implicit repeals of previous laws will be disfavored; that statutes will not ordinarily be read to interfere with the president's power in foreign affairs.

4. *Substantive purposes.*—Finally, interpretive norms may serve substantive purposes wholly apart from statutory meaning, interpretive intent, or the lawmaking process. The category of substantive principles spans a wide range. Substantive principles may reflect an objectionable judicial

10. 1 U.S.C. §§ 1–6.

value judgment; but they might derive instead from policies that are easy to defend.

Interpretive norms often derive from constitutional principles. It is a familiar idea that when a statute might be interpreted to be valid or invalid, courts should construe it so that it survives constitutional challenge. There are numerous other substantive norms as well; many of them are without constitutional inspiration. Such norms counsel courts not to infer private causes of action from regulatory statutes, to assume that legislation of Congress applies only within the territorial jurisdiction of the United States, not to infer exemptions from taxation, to assume that criminal statutes require mens rea, to avoid irrationality, to counteract obsolescence, and (an outmoded norm) to protect common law rights in the absence of a clear statement from Congress.[11]

An Alternative Method

All of these considerations suggest an alternative to the conventional understandings of statutory interpretation. The statutory text is of course the starting point, but it becomes intelligible only because of background norms that give it content. Usually, those norms are so widely shared, and so uncontroversial, that the text "alone" appears to be a sufficient basis for interpretation; statements of this sort mean that together with generally agreed upon interpretive principles, the meaning of the text is not in doubt. In many cases, however, the text will produce ambiguity, overinclusiveness, or underinclusiveness.

In such cases, courts often must resort to background norms of various sorts. Such norms are sometimes designed to reveal meaning in the particular cases; they are sometimes intended to serve procedural or substantive policies. Sometimes they serve their intended purpose well, and those purposes are not terribly controversial. But sometimes the relevant purposes are disputable—consider norms calling for the protection of federalism, the rule of lenity in criminal cases, the protection of Indian tribes—and sometimes it is unclear that those purposes are in fact well served by the particular principle.

Interpretive Principles for the Regulatory State

Background norms vary from one country to another; the relevant principles diverge sharply in accordance with the political understandings that prevail in the nation's legal culture. It is in this light that one should understand disagreements about interpretive norms in different legal systems. Background norms also differ from one period to another within any particular country. In the United States, for example, the principles

11. See, e.g., United States v. Wells Fargo Bank, U.S. (1988) (no implied exemptions from taxation); Foley Bros. v. Filardo, 69 Sup. Ct. 575 (1949) (laws apply only within the territory of the United States).

of the late nineteenth and early twentieth centuries reflected a belief in common law ordering that was largely repudiated in the 1930s.

In this section, I provide an outline of the basis for and content of interpretive norms that are designed to promote constitutional purposes, and the basic goals of deliberative government, in the post–New Deal period.

Sources of interpretive principles

a) In general.—Interpretive norms require sources; and it is important to be quite explicit about them here. The first and most straightforward is the Constitution. Understandings about constitutional arrangements provide a significant amount of the background against which statutory construction occurs. To be sure, there is frequently uncertainty about constitutional meaning; and statutory disputes are sometimes, in part, a function of a broader dispute about the nature of the constitutional structure. Often, however, there is sufficient consensus about the Constitution to ground a large part of interpretive practice.

The second source of interpretive norms—also straightforward—consists of understandings about the appropriate and actual performance of governmental institutions. One might, for example, conclude that appropriations statutes are written hastily and without deliberation; that in the face of ambiguity, courts should defer to administrative interpretations of law; or that, on the contrary, regulatory agencies should have little discretion. If one were simply describing statutory construction, one would find a number of background norms traceable to understandings of precisely this sort.

b) The regulatory state.—The third and final source of background norms involves the intended functions of regulatory statutes and the ways in which such statutes tend to fail in practice. A detailed discussion of these points would be well beyond the scope of the present discussion, but a brief outline will be useful.

The goals of regulatory regimes span a wide range. Some statutes are designed to promote economic efficiency, others to redistribute resources in public-regarding ways, to reflect social values or aspirations, to counteract the social subordination of disadvantaged groups, or to enforce the terms of pure interest-group deals. Moreover, statutes fail for a variety of reasons. Sometimes they are adversely affected by interest-group power, collective action problems, the absence of political accountability or deliberation, obsolescence, or lack of coordination with other statutes. Other statutes fail because of implementation problems. Statutes designed to redistribute resources, for example, may fail to take account of the way that the market can nullify the intended effects. Statutes designed to reduce or eliminate the social subordination of disadvantaged groups, or to promote noncommodity values like environmental diversity, are frequently subject in the implementation process

to precisely the same obstacles that make such statutes necessary in the first place. Government failure thus replicates market failure.[12]

Conclusions of this sort are principally of interest to regulators in the legislative and executive branches; but they are relevant to the judiciary as well. Legal interpretation of regulatory statutes has in fact been responsive to understandings about the likely function and failure of those statutes. Moreover, legal interpretation is likely to be improved by such understandings. It is not simple, to be sure, to make such understandings operational—to explain exactly how they should be used in hard cases. But it is not difficult to show how understandings of regulatory function and failure have worked, and should work, to produce norms—or so I will be arguing.

The Principles

In this section, I briefly describe interpretive norms drawn from the three sources described above.

1. *Constitutional norms.*—The constitutional norms that play a role in statutory interpretation span a wide range. The Constitution provides the backdrop against which statutes are written and interpreted. Moreover, federal courts underenforce many constitutional norms, and for good reasons.[13] Institutional constraints—most notably, the courts' limited fact-finding ability and weak electoral legitimation—lead courts to be reluctant to vindicate constitutional principles with the vigor appropriate to governmental bodies with a better democratic pedigree. It follows that some statutes should be understood as responses to Congress's constitutional responsibilities even if courts would not require Congress to carry out those responsibilities in the first instance. Relatively aggressive statutory interpretation—pushing statutes away from constitutionally troublesome ground—provides a way for courts to vindicate norms that do in fact have constitutional status.

a) *Federalism.*—In the system of American public law, the basic assumption is that states have authority to regulate their own citizens and territory. This assumption justifies an interpretive norm calling for a clear statement in order to support federal preemption of state law.

b) *Political deliberation; the constitutional antipathy to naked interest-group transfers.*—The American constitutional regime is built on hostility to measures that impose burdens or grant benefits merely because of the political power of private groups; some public value is required for governmental action. This norm, traceable to the origins of the Constitution and firmly rooted in current law, has a number of implications for statutory interpretation. It suggests, for example, that courts should develop in-

12. See, generally, C. Sunstein, *After the Rights Revolution: Reconceiving the Regulatory State* (Cambridge, Mass.: Harvard University Press, 1990), in press, for citations and details.

13. See L. Sager, "Fair Measure: The Legal Status of Underenforced Constitutional Norms," *Harvard Law Review* 91 (1978): 1212.

terpretive strategies to promote deliberation in government—by, for example, remanding issues involving constitutionally sensitive interests or groups for reconsideration by the legislature or by regulatory agencies when deliberation appears to have been absent. It also suggests that statutes that embody mere interest-group deals should be narrowly construed. Much of American administrative law is founded on these ideas.[14]

c) *Political accountability; the nondelegation principle.*—In the period since the New Deal, courts have permitted Congress to delegate exceptionally broad policy-making authority to regulatory agencies. Such broad delegations were at one time thought to offend Article I of the Constitution, which vests legislative power in Congress. In recent years, courts have been reluctant to enforce the nondelegation doctrine. But through statutory construction, courts are sometimes able to vindicate the constitutional principle against delegation of legislative power. They can do so, for example, by narrowly construing delegations of policy-making power.

d) *Property and contract rights.*—In the aftermath of the New Deal reformation, courts have been reluctant to use the Constitution's explicit protection of property and contract in a way that would interfere with social and economic regulation. But courts can vindicate those rights less intrusively through the narrow construction of regulatory statutes that raise serious constitutional doubts under the contracts and takings clause.

e) *Welfare rights.*—In recent years, the Supreme Court has rejected claims that the Constitution guarantees rights of basic subsistence; the resistance stems largely from reasons of democratic legitimacy. But perhaps a Court that saw substantive force in the claim for constitutional welfare rights would attempt to vindicate that claim, not through the Constitution itself, but through aggressive statutory construction so as to ensure against irrational or arbitrary deprivations of benefits. Such an approach would tend to produce even-handedness in the distribution of welfare in a democracy that has committed itself to a "social safety net." Aggressive statutory construction, requiring a clear statement for exclusions, might produce many of the advantages of recognition of constitutional welfare rights without imposing nearly so severe a strain on the judiciary.

f) *Disadvantaged groups.*—Some of the most difficult constitutional questions involve the extent of protection afforded to disadvantaged groups by the equal protection clause; and in many cases, courts have afforded less protection than they might because of a desire to minimize interference with the political branches of government. Thus the Court has required a showing of discriminatory "intent" in order to make out a violation of the equality principle, and in various ways it has limited its protection of blacks, women, and the handicapped.[15] In this light, statutes that provide protection for such groups, or against discriminatory

14. See C. Sunstein, "Interest Groups in American Public Law," *Stanford Law Review* 38 (1985): 29.

15. See Washington v. Davis, 426 U.S. 229 (1976); Personnel Admin. v. Feeney, 442 U.S. 256 (1979); City of Cleburne v. Cleburne Living Center, 473 U.S. 432 (1985).

effects, might well be seen as a legislative response to its (judicially underenforced) constitutional duties. Through statutory constructions, courts can ensure that the relevant norms are vindicated.

2. *Institutional concerns.*—A number of interpretive principles respond straightforwardly to institutional concerns, having to do with the actual or appropriate functioning of governmental entities. Most of these principles are straightforward and can be discussed quite briefly.

a) *Appropriations statutes.*—Courts construe appropriations provisions quite narrowly. That principle can be connected with judicial understandings about the character of the appropriations process, in which legislative deliberation is highly unlikely, and interest-group power is peculiarly likely to make its effects felt. Narrow construction of appropriations statutes promotes deliberation in government.

b) *The presumption in favor of judicial review.*—Courts presume that the legislature has not precluded judicial review of agency decisions. This presumption is partly attributable to a belief—vindicated by recent experience[16]—that regulatory agencies are susceptible to factionalism and self-interested representation. Judicial review operates as a powerful ex ante deterrent against dangers of this sort; it is also an ex post corrective. If Congress is to eliminate judicial review, it must do so unambiguously.

c) *Implied exemptions from taxation.*—Courts do not infer exemptions from taxation. This principle is partly a result of a desire to protect the Treasury; it is partly a product of a view that exemptions from taxation frequently reflect the influence of well-organized private groups.

d) *The rule of law.*—Courts interpret statutes so as to promote the rule of law. The most conspicuous example is the rule of lenity: criminal statutes will be construed favorably to the defendant, above all to promote fair notice. Other illustrations include a judicial assumption that statutes do not apply retroactively, that administrative discretion is narrow, and that administrators have the authority to proceed by rule.

3. *Counteracting statutory failure.*—I have suggested that a number of interpretive norms might be designed to counteract failures in social and economic regulation. In this space, it will be possible only to list the relevant norms. Where there is interpretive doubt, a perception of characteristic pathologies in regulatory legislation helps in statutory construction. Courts should, I suggest, construe regulatory statutes so that (1) politically unaccountable actors are prohibited from deciding important issues; (2) collective action problems do not subvert statutory programs; (3) different regulatory statutes are, to the extent possible, coordinated into a coherent whole; (4) obsolete statutes are kept consistent with changing

16. See, e.g., K. Schlotzman and J. Tierney, *Organized Interests and American Democracy* (New York: Harper & Row, 1986); Stewart, "The Reformation of American Administrative Law," *Harvard Law Review* 88 (1975): 1193.

developments of law, policy, and fact; (5) procedural qualifications of substantive rights are kept narrow; (6) the complex systemic effects of regulation are taken into account; and, most generally, (7) irrationality, measured against the statute's own purposes, is avoided. These principles apply across the universe of regulatory statutes, though their application will vary with the statutory function.

In light of the conventional difficulties in the implementation process, courts should also, I suggest, generously construe statutes designed to protect traditionally disadvantaged groups and noncommodity values. Courts should attempt to ensure as well that the costs of regulation are proportionate to the benefits and that de minimis exceptions to regulatory requirements are permitted or required. Finally, statutes that embody interest group transfers should be narrowly interpreted.

It should be clear that some of these ideas are a product of contestable substantive theories. Some of them depend for their support on views about the usual nature of regulatory failure; and it would of course be possible to accept some of the principles while rejecting others. Notwithstanding their number and variety, moreover, the various norms are united by certain general goals. These include, above all, the effort to promote deliberation in government, to furnish surrogates for it when it is absent, to limit factionalism and self-interested representation, and to help bring about political equality. All of the norms are responsive to an understanding of the various reasons for statutes and of the various ways that they tend to fail in practice. Of course it would be necessary to develop rules of priority and harmonization in cases of conflict.

Illustrations

1. *The Delaney Clause.*—I have referred to the Delaney Clause, which prohibits the sale of food additives that "induce cancer when ingested by man or animal" or are found "after tests which are appropriate for the evaluation of the safety of food additives, to induce cancer in man and animal." As we have seen, the drafters of the clause believed that few additives caused cancer and that those that did so were extremely dangerous. By the 1980s, however, it was clear that many substances were carcinogenic, but that a number of them created exceptionally minor risks. These substances included natural dietary constituents posing dangers that surpass those of artificial chemicals.

These developments severely undermined the assumptions of the Congress that enacted the Delaney Clause, and indeed made it appear quite perverse in many of its applications—banning as it did substances that posed little or essentially no risk to health. One consequence was that manufacturers, having been forbidden to use food additives posing a de minimis risk of cancer, used instead additives that caused far more serious risks of other diseases—thus creating what is quite possibly a significant increase in illnesses and deaths from food additives. In response, the Food and Drug Administration (FDA)—invoking background prin-

ciples against obsolescence and in favor of de minimis exceptions—interpreted the clause to permit exemptions of trivial risks. It said that it would approve food additives that posed a de minimis risk of cancer.

The United States Court of Appeals for the District of Columbia held that the FDA's position was unlawful.[17] In the court's view, the Delaney Clause was simply unambiguous on the point. But this interpretation was far from inevitable. Read in its context, there was no such clarity. Congress simply did not focus on the question of de minimis risks. It is far from clear that the court's mechanical reading of the statute was sensible from the perspective of its original authors, whose assumptions simply did not hold in the 1980s. The background against which the Delaney Clause was written was so different from present circumstances that the statutory terms "induce cancer" must be treated as ambiguous. Whether a substance "induces" cancer within the meaning of the clause might well be a function of the degree of risk that it posed. In these cirumstances, interpretation of the clause to permit de minimis exceptions seems quite consistent with permissible understandings of judicial interpretation and quite sensible to boot. The court's decision to the contrary was therefore misguided.

2. *Pennhurst.*—In *Pennhurst State School and Hospital v. Halderman,*[18] a group of mentally retarded people brought suit against a facility that was found to contain dangerous conditions. Many of the residents had been physically abused, brutally mistreated, or drugged; and the facility was utterly inadequate for the habilitation of the retarded. As the basis of their legal claim, the plaintiffs invoked the "bill of rights" in the Developmentally Disabled Assistance and Bill of Rights Act.[19] The bill of rights contained a set of legislative findings, including the following:

> 1. Persons with developmental disabilities have a right to appropriate treatment, services, and habilitation for such disabilities.
>
> 2. The treatment, services, and habilitation for a person with developmental disabilities should be designed to maximize the developmental potential of the person and should be provided in a setting that is least restrictive of the person's personal liberty.
>
> 3. The Federal Government and the States both have an obligation to assure that public funds are not provided to any institution . . . that does not provide treatment, services, and habilitation which is appropriate to the needs of such persons.

The question in *Pennhurst* was whether the bill of rights created legally enforceable rights or whether it instead set out a set of goals or aspirations that could not be vindicated in court. Whether the bill of rights was legally enforceable was an uncertain issue in light of the text

17. See Public Citizen v. Young, 831 F.2d 1108 (D.C. Cir. 1987).
18. Pennhurst State School and Hospital v. Halderman, 451 U.S. 1 (1981).
19. Developmentally Disabled Assistance and Bill of Rights Act, 42 U.S.C. 6010.

and history of the act. Without a background interpretive principle, the case was probably impossible to resolve.

In holding for the governmental defendant, the Supreme Court did not deny that the conventional sources of interpretation left the case in equipoise. Instead it invoked a background norm: "If Congress intends to impose a condition on the grant of federal moneys, it must do so unambiguously. By insisting that Congress speak with a clear voice, we enable the States to exercise their choice knowingly, cognizant of the consequences of their participation." *Pennhurst* was thus decided as a result of an interpretive norm that was a result of the Court's understanding of the constitutional background—a norm that could not be connected to the statute at issue in the case.

For reasons explored above, federalism principles are properly invoked, at least in ordinary settings, to require a clear statement from Congress for the imposition of significant duties on the states. If the bill of rights had been held legally enforceable, the states would have faced an enormous financial burden, one that they had not, in view of the ambiguity of the statute, necessarily agreed to incur. But two considerations suggest that *Pennhurst* was incorrectly decided. First, federalism principles have much less force in cases in which Congress is attempting to protect a traditionally disadvantaged group from the political processes of the states. Here the ordinary presumption in favor of state autonomy is countered by the history and implementation of the fourteenth amendment—a self-conscious limitation on state power. Invocation of principles of state autonomy in the context of a socially subordinated group—to justify a narrow reading of a statute enacted on its behalf—is positively perverse in light of constitutional structure and history.

Second, an underenforced constitutional norm calls for aggressive construction of statutes involving the developmentally disabled. The Court has been extremely cautious in using the Constitution to protect the disabled, largely for institutional reasons. In holding that the equal protection clause does not entitle the mentally retarded to special protection, for example, the Court stressed that federal and state legislatures have in fact responded to the pervasive mistreatment of the retarded, and that courts should be reluctant to intrude in so dramatic a way into democratic processes.[20] In these circumstances, it remains possible to argue that the best substantive theory of the equal protection clause in fact accords special protection to the disabled, and that when other branches give such protection, they are carrying out their constitutional responsibilities.

It follows that when Congress has taken steps to provide safeguards for that group, the Court should take into account the underenforced character of the equality norm in order to give statutes involving the disabled a hospitable rather than grudging interpretation. It is puzzling

20. See City of Cleburne v. Cleburne Living Center, 105 U.S. 3249 (1985).

indeed that the *Pennhurst* Court relied on a background norm in favor of federalism in a case involving a group and a right safeguarded by the fourteenth amendment.

CONCLUSION

Because language "by itself" is without meaning, and in light of the existence of gaps or ambiguities in hard cases, interpretive norms of various sorts are indispensable—desirable and in any case inevitable. Statutory interpretation cannot go forward without background principles that orient reader to text, fill gaps in the face of legislative silence, or provide the backdrop against which to read linguistic commands. Some such principles aid courts in discerning the meaning of particular statutes, or help to implement Congress's actual or likely interpretive instructions; others are rooted in constitutional concerns; others are based on assessments of institutional performance; and still others attempt to respond to characteristic failings of regulatory legislation. In all of these respects, norms of statutory construction, far from being obsolete or representing an illegitimate judicial value judgment, must occupy a prominent place in the theory and practice of statutory interpretation.

The ultimate task is to develop a set of interpretive norms or background understandings—sensitive to constitutional structure, institutional design, the New Deal reformation, and the diverse functions and failings of governmental actors and statutory regimes—with which to approach social and economic regulation in a system that has largely abandoned common law categories. And it is possible to generate a series of norms, all with support in current law, that can help accomplish some of the goals of deliberative government in the post–New Deal period. In this way, statutory construction can serve as an ally of other, more ambitious strategies designed to promote some of the original constitutional goals in a dramatically changed legal environment. It is far too much to expect statutory construction to respond to all of the failings of the regulatory state. But it is not too much to expect that the process of interpretation can make the situation better rather than worse.