9-2000

Rules and Judicial Review

Emily Sherwin
Cornell Law School, els36@cornell.edu

Legal Theory, 6 (2000), 299–321. Printed in the United States of America

RULES AND JUDICIAL REVIEW

Emily Sherwin*
University of San Diego

Judicial review of statutes on constitutional grounds is affected by a cluster of doctrinal practices that are generally accepted, but not very well explained, by the courts and not entirely consistent with each other. Courts usually judge statutes "as applied" rather than as written;[1] they favor "severance" of valid applications of statutes from invalid or possibly invalid applications when possible;[2] and they interpret statutes in ways that avoid constitutional difficulty.[3] These overlapping practices presumably are intended to preserve legislation, and hence are associated with a modest conception of the role of courts in government. Yet they are not always modest in operation.[4]

The objective of this article is to examine the effect of statute-saving devices such as as-applied adjudication, severance, and narrowing interpretation, and to consider whether a rule-oriented analysis of judicial review can offer any insights about how courts should handle unconstitutional statutes.[5]

I. AS-APPLIED CHALLENGES, SEVERABILITY, AND SAVING INTERPRETATIONS

A. Summary of Current Practice

One frequently stated principle of judicial review is that constitutional challenges to statutes normally will be treated as challenges to statutes as

*Professor of Law, University of San Diego. Thanks to Larry Alexander and Kevin Clermont for helpful comments.

1. *See generally* Michael C. Dorf, *Facial Challenges to State and Federal Statutes*, 46 STAN. L. REV. 235 (1994); Marc E. Isserles, *Overcoming Overbreadth: Facial Challenges and the Valid Rule Requirement*, 48 AM. U. L. REV. 359 (1998).

2. *See generally* Mark Movsesian, *Severability in Statutes and Contracts*, 30 GA. L. REV. 41 (1995); John Copeland Nagle, *Severability*, 72 N.C. L. REV. 203 (1993); Robert Stern, *Separability and Separability Clauses in the Supreme Court*, 51 HARV. L. REV. 76, 82–106 (1937).

3. *See generally* Frederick Schauer, Ashwander *Revisited*, 1995 SUP. CT. REV. 71; Adrian Vermeule, *Saving Constructions*, 85 GEO. L.J. 1945 (1997).

4. Several commentators have equated the practices mentioned in the text to judicial revision of statutes. *See* Dorf, *supra* note 1, at 292–93 (severance results in a "judicially rewritten law"); Nagle, *supra* note 2, at 220 (the product of severance is "akin to a new statute"); Schauer, *supra* note 3, at 80–81 (narrowing construction is a form of "redrafting").

5. For a comprehensive discussion and defense of the essential connection between constitutional rights and rules, to which I am much indebted, see Matthew D. Adler, *Rights Against Rules: The Moral Structure of American Constitutional Law*, 97 MICH. L. REV. 1 (1998).

applied in particular cases, rather than as challenges to statutes as they appear on the books. The Supreme Court, for example, has stated that to succeed in a facial challenge to a statute, a litigant must "establish that no set of circumstances exists under which the act would be valid."[6] Only if the statute has no valid, independent applications, or if the case falls within the special "overbreadth" and "vagueness" doctrines associated with freedom of expression and perhaps with other liberties, will the court hold the statute wholly invalid, on its face.[7] The usual consequence of as-applied adjudication is that if the court finds the statute valid as applied to the challenger, the challenger loses and the statute remains in force, even if other applications might be impermissible.[8]

A second principle of judicial review holds that unconstitutional provisions of statutes normally will be "severed" from other, valid provisions. If unconstitutional (or constitutionally questionable) provisions are found to be severable from other, valid provisions, the court will enforce the valid portion of the statute.[9] If the court determines that the statute is not severable, however, it will address the statute as a whole, and proof that some provisions are invalid means that the statute is unenforceable in its

6. United States v. Salerno, 481 U.S. 739, 745 (1987). Some have read the statement quoted in the text to mean that a statute cannot be challenged on its face if it has some constitutional applications. *See* Dorf, *supra* note 1, at 250. Marc Isserles interprets it to mean that a statute is facially invalid if the terms of the statute, "measured against the relevant constitutional doctrine, and independent of the constitutionality of particular applications," state a constitutionally invalid rule. Isserles, *supra* note 1, at 387.

7. An example of a statute with no valid applications is one that contains a discriminatory predicate. For example, a statute that prohibits assaults by blacks is invalid in all applications, although assaults can be prohibited, including assaults by blacks. *See* Lawrence A. Alexander, *Is there an Overbreadth Doctrine?*, 22 SAN DIEGO L. REV. 541, 545 (1985) (giving examples of statutes that are necessarily invalid in all applications because their predicates are underinclusive). *See also* R.A.V. v. City of St. Paul, 505 U.S. 377, 381 (1992) (hate speech ordinance facially invalid because it targeted the content of speech); Isserles, *supra* note 1, at 386–95 (discussing facial challenges based on a constitutional violation inherent in the terms of the statute).

On First Amendment overbreadth and vagueness doctrines, see generally Thornhill v. Alabama, 310 U.S. 88 (1940); Lawrence H. Tribe, AMERICAN CONSTITUTIONAL LAW 1022–24, 1033–35 (2d ed. 1988); Alexander, *supra;* Richard H. Fallon, Jr., *Making Sense of Overbreadth*, 100 YALE L.J. 853, 854 (1991); Alfred Hill, *The Puzzling First Amendment Overbreadth Doctrine*, 25 HOFSTRA L. REV. 1063, 1083 (1997); Henry Paul Monaghan, *Overbreadth*, 1981 SUP. CT. REV. 1. First Amendment Overbreadth is briefly discussed at notes 37–39, *infra*.

8. *See, e.g.*, Yazoo & Mississippi Valley R.R. v. Jackson Vinegar Co., 226 U.S. 217 (1912) (railroad that refused to settle a nonfrivolous damage claim could not challenge a railroad liability statute that might be unconstitutional when applied to frivolous claims).

9. *See, e.g.*, INS v. Chadha, 462 U.S. 919, 931–35, 959 (1983); Alaska Airlines v. Brock, 480 U.S. 678, 684–97 (1987). John Nagle lists five ways in which a question of severability can arise: A litigant may claim that a statute is invalid in its entirety because unconstitutional provisions are not severable from the rest of the statute; a litigant may claim that one provision is unenforceable because it is not severable from other invalid provisions; a litigant may claim that one application of a statute is unacceptable because it is not severable from other invalid applications; a litigant may argue that provisions of a statute are not severable, and therefore another litigant cannot enforce any part of the statute; or a litigant may challenge a statute as underinclusive or overinclusive. Nagle, *supra* note 2, at 208–09.

entirety.[10] Thus, the effect of severance is often much the same as the effect of as-applied adjudication.[11]

Despite the similar consequences of severance and as-applied decision making, a somewhat different set of rules applies to severability. The traditional formula asks (1) whether the remainder of the statute can operate coherently as an independent statute, and (2) whether the legislature would have enacted the remainder of the statute alone.[12] Although severability has not always been favored, courts now generally presume that statutes are severable, stating that legislation should be preserved to the greatest extent possible or that severance is a way to avoid unnecessary constitutional decisions.[13] Statutes often address the question of severability by providing that in case of challenge their valid provisions should be preserved, but courts have not treated these severability clauses as decisive. Inclusion of a severability clause in the statute at best raises or reinforces the presumption of severability.[14]

A third principle affecting judicial review of statutes is that courts should interpret statutes in ways that avoid unconstitutionality, or that avoid constitutional questions.[15]

10. *See, e.g.*, Planned Parenthood of Central Missouri v. Danforth, 428 U.S. 52, 81–84 (1976).

11. Michael Dorf has argued that the normal practice of judging the constitutionality of statutes as-applied is in effect a practice of severance. Dorf, *supra* note 1, at 249–50. *See also* Stern, *supra* note 2, at 82–87 (discussing separable applications of statutes); Dorf equates severability and as-applied adjudication.

With severability in mind, First Amendment overbreadth doctrine can be conceived as an exception to the normal practice of severing constitutional from unconstitutional applications of statutes, in response to special concerns about the deterrent effect of the statute on protected expression. See Dorf, *supra* note 1, at 261–62 (equating a deterrent-based account of First Amendment overbreath with "a presumption of nonseverability where rights of expression are concerned").

12. *E.g.*, Champlin Refining Co. v. Corporation Comm'n of Oklahoma, 286 U.S. 210, 234 (1932), *overruled in other respects*, Philips Petroleum Co. v. Oklahoma, 340 U.S. 190 (1950) ("Unless it is evident that the Legislature would not have enacted those provisions which are within its power, independently of that which is not, the invalid part may be dropped if what is left is fully operative as a law.") ; New York v. U.S., 505 U.S. 144, 186 (1992) (quoting *Champlin, supra*); Alaska Airlines v. Brock, 480 U.S. 678, 684 (1987) (quoting *Champlin, supra)*; Dorf, *supra* note 1, at 285 (stating that forty-eight of fifty states follow this test); Nagle, *supra* note 2, at 210–211 (discussing the test).

13. *See* Regan v. Time, Inc., 468 U.S. 641, 652 (1984) ("[T]he presumption is in favor of severability."); Nagle, *supra* note 2, at 218–20 (tracing the history of presumptions for and against severability); *cf. Alaska Airlines*, 480 U.S. at 686 (stating that "In the absence of a severability clause, . . . Congress' silence is just that—silence," but also quoting the standard test, which calls for severance unless it is "evident" that legislature would not have enacted the valid portion of the statute alone).

14. *See, e.g., Alaska Airlines*, 480 U.S. at 686 (severability clause creates a presumption in favor of severability). On severability clauses, *see generally*, Movsesian, *supra* note 2, at 73–77; Nagle, *supra* note 2, at 234–46; Stern, *supra* note 2, at 120–28.

15. *See, e.g.*, De Bartolo Corp. v. Florida Gulf Coast Building & Construction Trades Counsel, 485 U.S. 568, 575 (1988) ("[W]here an otherwise acceptable construction of a statute would raise serious constitutional problems the Court will construe the statute to avoid such problems unless such construction is plainly contrary to the intent of Congress."); Ashwander v. Tennessee, 297 U.S. 288, 348 (1936) (Brandeis, J., concurring) (citing Crowell v. Benson, 285

This principle differs from as-applied adjudication or severance in that it purports to exonerate the statute completely from charges of unconstitutionality. If the court arrives at an interpretation that avoids constitutional difficulties simply because it believes that this is what the legislature intended, no question of as-applied decision making or severability can arise. It is likely, however, that a saving interpretation typically entails rejecting the most plausible interpretation of a statute in favor of a constitutionally safe, but less plausible, reading.[16] Otherwise, the principle that courts should read statutes in ways that avoid constitutional questions adds nothing to interpretation. As a result, the effect of a saving construction is similar to that of a decision to uphold a statute as applied or to sever provisions of the statute: It alters the original legislation.[17] If the interpretation the court settles on is not the interpretation the legislature intended, but the best interpretation that can be reconciled with the Constitution, the court has either severed applications of the statute, or simply rewritten it.[18]

The common feature of the three practices just described is that in each case, the effect of the court's decision is to create a new rule. When a court finds an unconstitutional application to be severable from other valid applications, what remains after severance is a new, narrower law.[19] When a court

---

U.S. 22, 62 n.8 (1932) ("'When the validity of an act of the Congress is drawn in question, and even if a serious doubt of constitutionality is raised, it is a cardinal principle that this court will first ascertain whether a construction of the statute is fairly possible by which the question may be avoided.'").

16. Frederick Schauer has made this argument persuasively. Schauer, *supra* note 3, at 83–84. It can be argued that legislatures intend statutes to be given a meaning that conforms to constitutional requirements. *Cf.* Alexander, *supra* note 7, at 542–43 (arguing that constitutional requirements can be read into statutes either as external limits or "[b]y indulging the fiction that the legislature always intends to incorporate the Constitution into each of its laws). Schauer points out, however, that our political system does not impose sanctions on legislators who overreach constitutional boundaries; hence there is no reason to assume, as a matter of intent, that statutes incorporate constitutional limits. Schauer, *supra* at 92–93.

17. *See* Schauer, *supra* note 3, at 80–81; Stern, *supra* note 2, at 95.

18. For strong criticism of the practice of avoiding constitutional questions by narrow construction of statutes, *see* Schauer, *supra* note 3, at 91–98.

Adrian Vermeule presents an interesting analysis of the relation between severability and the rule that courts should construe statutes to conform to constitutional requirements. According to Vermeule, an early version of this rule stated that, given two possible interpretations, one constitutional and one unconstitutional, the court should choose the constitutional interpretation. In the version now prevalent, however, the rule states that courts should interpret statutes to *avoid* constitutional questions. Thus, the early form of the rule required a constitutional decision (namely, that one reading of the statute was in fact unconstitutional), and therefore operated much like a severability decision following a finding of unconstitutionality. In comparison, the current form of saving construction deflects constitutional questions. Vermeule, *supra* note 3, at 1948–49. In Vermeule's view, this early form of saving construction, as well as the typical form of severance following a finding of unconstitutionality, are more legislation-friendly practices because a saving construction to avoid constitutional questions forecloses applications of the statute that might be valid. *Id.* at 1960.

19. *See* Adler, *supra* note 5, at 125–32 (comparing facial invalidation of statutes to various forms of "optimal revision"); Dorf, *supra* note 1, at 292–93 (discussing severance); Nagle, *supra* note 2, at 220 (same).

If the court defers constitutional decision, by presuming that other, possibly invalid, provisions are severable or by simply holding that the statute is acceptable as applied (and therefore

interprets a statute in a way it would not endorse in order to avoid constitutional difficulty, the result is a new, narrower law.[20]

It might be thought that as-applied adjudication involves a different type of judicial action—that here the court is not altering the rule of the statute but only recognizing that the particular acts of the challenger are constitutionally privileged.[21] To view as-applied adjudication in this way, however, one must conceive of the Constitution as a body of law directly applicable to actors, rather than as a set of constraints on law. Other participants in this symposium have argued convincingly that this conception is neither textually sound, logically appealing, nor borne out by the practice of the Supreme Court. Instead, a constitutional right is best conceived of not as a right to act, but as a right to be governed and judged by valid laws.[22] If one accepts this premise, a decision upholding or invalidating a statute as applied is not a decision about privilege, but a decision that part of the statute—a severable subrule applied to the challenger—is valid. Accordingly, the result of the decision is, once again, a new, narrower law.

Thus, all three of the practices I have discussed—as-applied adjudication, severance, and avoidance of constitutional questions through narrow interpretation of statutes—result in judicial promulgation of new law. In the interest of preserving legislation, the court is legislating.

B. Methodology

Commentary on severability and as-applied adjudication suggests at least four methods courts might use to determine when to save statutes by separating valid from invalid provisions or applications. For this purpose, I

---

implicitly severable), it does not immediately amend the statute, but it sets the stage for an amendment. In the case of Supreme Court review of a state statute, it is the state court that ultimately amends the statute, if it chooses to find the statute severable. *See*, *e.g.*, Yazoo & Mississippi Valley R.R. v. Jackson Vinegar Co., 226 U.S. 217, 219–20 (1912) (upholding statute as applied and leaving further applications to the state); Dorf, *supra* at 283–84 (discussing federal review of state statutes); Monaghan, *supra* note 2, at 6–8 (discussing *Yazoo*, severability, and state courts).

20. *See* Schauer, *supra* note 3, at 80–81 (narrowing interpretation is a form of "redrafting"); Stern, *supra* note 2, at 82 ("the result can be described as amending the law").

21. *See* Adler, *supra* note 5, at 5–7 (describing a "Direct Account" of constitutional rights, in which a reviewing court holds the treatment of a particular litigant to be unconstitutional, in contrast to a "Derivative Account," in which the court invalidates or revises the rule under which the litigant was sanctioned); Dorf, *supra* note 1, at 244–46 (describing but ultimately rejecting a "privileged-conduct-only" view of constitutional rights); Monaghan, *supra* note 7, at 4–5 (describing an activity-oriented view of constitutional rights).

22. *See* Adler, *supra* note 5, at 91–132 (defending the "Derivative Account" of constitutional rights); Alexander, *supra* note 7, at 545 (suggesting that the Constitution limits the reasons for which government can act, rather than protecting particular acts); Dorf, *supra* note 1, at 242–50 (arguing that constitutional rights are best conceived as rights to be judged by a valid rule); Monaghan, *supra* note 7, at 4–8 (arguing that litigants have a right to be judged by a constitutionally valid rule).

shall assume that severability and as-applied adjudication are variants of the same problem.[23]

Perhaps the most obvious approach is to try to ascertain what the legislature actually intended—what did it want courts to do in case a part of the statute failed? As a general matter, the primary source of information about legislative intent, at least according to current Supreme Court practice, is statutory text. Beyond the text, courts refer to such standard tools as structure, purpose, and legislative history.[24] Some have suggested that the same methods should be applied to decide questions of severability.[25] One consequence is that a severability clause normally would be conclusive.[26]

An intent-based, primarily textual approach to severability and related questions is based on the sensible premise that severability is a matter of statutory interpretation—part of the meaning of the statute. At least in ordinary severability cases outside the sphere of the First Amendment,[27] courts have no source for conclusions about severability apart from the statute itself. The difficulty with an approach based on actual intent is that severability questions are triggered by unplanned statutory failures. There may be cases in which a legislature anticipates a specific constitutional challenge and provides for severability, but a legislature that believes its statute to be constitutionally sound typically will have no intent on the question what to do if part of the statute is held invalid. This means both that severability clauses are included without particular instances of severance in mind,[28] and that structure, purpose, and legislative history will have little to say about actual legislative intent.[29] There simply is no actual intent about the problem before the court.

A second approach is a variant on the intent-based approach that seeks hypothetical, rather than actual, legislative intent. Based on the other intentions the legislature appears to have had in connection with the statute, what would it want a court to do when part of the statute has

23. *See supra* text following note 23.

24. I do not intend to take a position on the difficult problem of interpretation. *See generally*, LAW AND INTERPRETATION: ESSAYS IN LEGAL PHILOSOPHY (Andrei Marmor ed., 1995); William N. Eskridge, Jr. & Philip P. Frickey, *Statutory Interpretation as Practical Reasoning*, 42 STAN. L. REV. 321 (1990). It is enough for my purposes that courts search for evidence of intent in the text, and secondarily in such sources as structure, purpose, and legislative history. *See, e.g.*, Huffman v. Western Nuclear, Inc., 486 U.S. 663, 672 (1988) (quoting United States v. American Trucking Ass'ns, 310 U.S. 534, 543 (1940)) (text); INS v. Cardozo-Fonseca, 480 U.S. 421, 449 (1987) ("plain language . . . , symmetry . . . , and legislative history"); Nagle, *supra* note 2, at 232.

25. *See* Movsesian, *supra* note 2, at 73–82, proposing a textual approach, together with a "default rule" in favor of severability; Nagle, *supra* note 2, at 232–34 (proposing that courts treat severability as a question of intent, relying primarily on text).

26. *See* Movsesian, *supra* note 2, at 74–75; Nagle, *supra* note 2, at 234–46 (allowing an exception for "absurd results").

27. *See* note 27, *supra*, text at notes 37–39, *infra*.

28. *See, e.g.*, Stern, *supra* note 2, at 122–23 (commenting on indiscriminate use of separability clauses); *but cf.* Nagle, *supra* note 2, at 241 (tracing the history of a 1989 child-care bill, in which Senators anticipated particular challenges).

29. *See, e.g.*, Nagle, *supra* note 2, at 246–50.

failed?[30] Another way to frame the question is to examine the provisions that would remain after severance, consult legislative history, and ask whether the legislature would have enacted the remaining provisions alone.[31] Because legislative intent is admitted to be hypothetical, severability clauses are now evidence of intent, but by no means conclusive.[32]

This hypothetical intent approach is part of the test for severability now used by the Supreme Court and most other courts: To find a statute severable, the court must be satisfied that the remaining parts would have been enacted by the legislature, apart from those that have been excised.[33] The difficulty here is that hypothetical intent is quite speculative.[34] It can perhaps be given some content by reference to legislative purpose, or to the "enterprise" to which the statute belongs.[35] But it can easily deteriorate into a question of what ought to happen, in which case "legislative intent" adds nothing.

A third possibility is to concede that legislative intent on the question of severability is elusive or nonexistent, and to decide on other grounds. Under the prevailing test of severability, for example, the question whether remaining provisions can stand as a coherent statute can be viewed as a question independent of legislative intent.[36] Another example can be found in the special treatment courts give to statutes regulating expression. When statutes of this kind are "substantially overbroad" in the sense that they forbid a significant amount of constitutionally protected expression, courts refuse to treat the statutes as severable, and permit challengers whose own activities are not protected to challenge the statutes in their entirety.[37] The reason for this is that overbroad statutes are thought to deter the

30. This test was proposed by Robert Stern in 1937. Stern, *supra* note 2, at 98.

31. *See, e.g.*, Alaska Airlines v. Brock, 480 U.S. 678, 684, 685 (1987); Warren v. Mayor of Charlestown, 68 Mass. (2 Gray) 84 (1854).

32. *See, e.g.*, *Alaska Airlines*, 480 U.S. at 686 (severability clause raises presumption of severability); Stern, *supra* note 2, at 114–18, 121–25 (discussing severability clauses).

33. *See* materials cited *supra* note 12.

34. *See* Nagle, *supra* note 2, at 230. Mark Movsesian has argued that an approach based on hypothetical intent equates interpretation of statutes, which address and bind a large number of parties, with interpretation of contracts, which are two-party bargains. Movsesian, *supra* note 2, at 58–59, 66–73.

35. *See* Richard A. Posner, *Legal Formalism, Legal Realism, and the Interpretation of Statutes and Constitutions*, 37 CASE W. RES. L. REV. 179, 189–91 (1986) (conceiving interpretation as a process in which the judge acts as a subordinate faced with a failed communication, and asks what action will best carry out the "enterprise" in which the legislature is engaged).

36. *See, e.g.*, Champlin Refining Co. v. Corporation Comm'n of Oklahoma, 286 U.S. 210, 234 (1932) (invalid provisions can be severed "if what is left is fully operative as a law"); Nagle, *supra* note 2, at 215–16 (describing this as an "objective" component of the inquiry). Of course, the incoherence of the remaining provisions can be characterized as a reason to believe the legislature would not intend a severance, *see Alaska Airlines*, 480 U.S. at 684 ("Congress could not have intended [severance] if the balance of the legislation is incapable of functioning independently"), but incoherence rather than intent is doing most of the work.

37. *See, e.g.*, Thornhill v. Alabama, 310 U.S. 88 (1940) (facial review); Broaderick v. Oklahoma, 413 U.S. 601, 615 (1973) (substantial overbreadth). *See generally* Alexander, *supra* note 7; Fallon, *supra* note 7; Hill, *supra* note 7; Monaghan, *supra* note 7. At least, federal courts will not presume that state statutes are severable. *See* Dorf, *supra* note 1, at 283–85.

exercise of important rights. The practice of facial invalidation removes the immediate deterrent, and may also encourage legislatures to be especially careful when they regulate expression.[38] In any event, courts applying overbreadth doctrine are acting for constitutional or prophylactic reasons that are independent of the statute itself.[39] The obvious question raised by a non-intent-based approach to severability is, What should count as reason for or against severance?[40]

A final approach is to set aside particularized inquiry into either legislative intent or the desirability of severance, and instead establish a rule or presumption governing severability and as-applied adjudication. The Supreme Court's statements about as-applied decision making represent a strong presumption in favor of that practice;[41] its statements about severability are less clear but suggest that a favorable presumption is at work.[42] An alternative would be to establish a clear statement rule to the effect that courts will consider statutes fully severable unless the legislature provides otherwise.[43]

Commentators have suggested a number of reasons in support of a presumption favoring severability and as-applied adjudication. Severance of

38. *See, e.g.*, Fallon, *supra* note 7, at 867–70, 884–88 ("chilling effect"), 887 (incentives for legislatures).

39. On the distinction between constitutional and prophylactic versions of First Amendment overbreadth doctrine, *see* Fallon, *id.* at 867–75 (finding both constitutional and prophylactic bases for overbreadth). *See also* Alexander, *supra* note 7, at 552–54 (proposing a balance between governmental interests in particular forms of law and protection of First Amendment rights); Dorf, *supra* note 1, at 261–64 (characterizing overbreadth as one instance of a constitutional non-severability principle in cases of fundamental rights); Monaghan, *supra* note 7, at 8–23 (suggesting that overbreadth is not a special First Amendment doctrine, but an application of the general constitutional right to be judged by valid laws, in a context that demands that laws regulate by the least restrictive means).

40. In this article, I shall have little to say about constitutional reasons for refusing to sever statutes, such as First Amendment overbreadth. Instead, I shall suggest a further set of criteria for severance and similar practices, which are based on the fact that these practices produce new legal rules, and are independent of both constitutional considerations and legislative intent. For my purposes, overbreadth stands as a special constitutional reason—independent of the concerns I raise about the justification of rules—for refusing to sever or uphold a statute as applied. I assume that overbreadth trumps whatever conclusions may arise from the analysis of severability and related questions that I propose.

In any event, my analysis is focused primarily on judicial review within a single jurisdiction—that is, federal court review of federal statutes or state court review of state statutes—which is not likely to be much affected by overbreadth. Commentators on overbreadth assume that within a jurisdiction, the fact that a statute regulates expression does not prevent the court from adopting a narrowing interpretation. *See, e.g.*, Fallon, *supra* note 7, at 877 (the effect of a Supreme Court overbreadth decision is that "the statute cannot be enforced until the state's courts provide a narrowing construction"). To the extent that the overbreadth doctrine is based on the immediate deterrent effect of a particular statute on First Amendment rights, a narrowing interpretation eliminates the problem. To the extent the doctrine serves to provide legislatures with incentives to be cautious when enacting statutes that affect important constitutional rights, perhaps it should preclude narrowing interpretations, even within a single jurisdiction; but that is a problem for First Amendment theorists.

41. *See, e.g.*, United States v. Salerno, 481 U.S. 739, 745 (1987).

42. *See, e.g.*, Alaska Airlines v. Brock, 480 U.S. 678, 686–87 (1987) (finding "no need to resort to a presumption").

43. John Nagle makes this proposal. Nagle, *supra* note 2, at 253–58.

valid from invalid applications of statutes is said to honor legislative supremacy,[44] to avoid unnecessary constitutional decision making,[45] and to limit judicial decision making to concrete cases arising from live disputes.[46] Severance also avoids a cumbersome process of re-enactment and accompanying loss of social benefits, particularly when the legislature has enacted a statute in omnibus form.[47] The problem with a rule or presumption in favor of severability and as-applied adjudication is that it is very blunt. Unless more errors of all types will be avoided by a uniform outcome than by individualized decision making on the question of severability, a rule or presumption may be a mistake.

The following discussion assumes that although severability is a question of interpretation in the sense that authority over what should happen to a failed statute lies with the legislature, it is also a question on which a search for legislative intent—that is, intent about specific severability problems that arise in constitutional adjudication—will normally be fruitless.[48] The exception is the case in which a legislature did anticipate a particular challenge, and provided for it; then the structure of government dictates

44. *See* Dorf, *supra* note 1, at 292 (severability is supported by the principle "that courts will not disturb Congressional policies reflected in legislation except insofar as the constitution requires"); Movsesian, *supra* note 2, at 80 ("separation-of-powers concerns"); Nagle, *supra* note 2, at 251 (severability honors the principle that "a court should give effect to a statute to the maximum extent permitted by the Constitution"); Vermeule, *supra* note 3, at 1961 (severance is designed to "put into effect . . . as broad a range of applications as possible").

45. *See* Dorf, *supra* note 1, at 292 (severability can promote "judicial restraint"); Nagle, *supra* note 2, at 250–51 (severance avoids constitutional questions). *But see* Vermeule, *supra* note 3, at 1950–52 (distinguishing between "ius tertii severance," in which the court upholds the application of a statute to the case before it and avoids a decision on the constitutionality of other applications, and "severance proper," in which the court finds the statute to be invalid as applied but also decides that other potential applications are valid; in the latter case, the court does not avoid constitutional decision making).

46. *See* Chicago v. Morales, 119 S. Ct. 1849, 1867–69 (1999) (Scalia, J., dissenting) (arguing against facial invalidation of statutes); Dorf, *supra* note 1, at 246 (discussing possible rationales for a view of constitutional rights as privileging particular actions, but ultimately rejecting such a view).

47. *See* Movsesian, *supra* note 2, at 80–81 ("severability comports with present legislative practice); Nagle, *supra* note 2, at 251–52 (referring to "empirical realities of the legislative process").

48. I have referred at several places in the text to interpretation and to noninterpretive judicial rulemaking. Implicit in these references, and in my later discussion of the functions of rules, is a particular conception of interpretation that focuses on authors' intent. Very briefly, I understand the meaning of a statute or other rule to be the meaning the rulemaking authority intended the words of the rule to have. The text of the rule is good, but not conclusive, evidence of intent. It follows from this conception of meaning that when a court departs from the authority's intent—whether on the basis of the literal meaning of words, underlying purposes the rule was designed to promote, or independent moral ideals to which the words can (or cannot) be fit—the court is engaged in a noninterpretive practice, and has in effect assumed the position of a rulemaking authority. This view is expressed in much more detail in Larry Alexander & Emily Sherwin, PAST IMPERFECT: RULES, PRINCIPLES, AND DILEMMAS OF LAW ch. 4 (forthcoming) (manuscript on file with author). On the many problems connected to interpretation, see generally Kent Greenawalt, STATUTORY INTERPRETATION: 20 QUESTIONS (1999); Keith E. Whittington, CONSTITUTIONAL INTERPRETATION: TEXTUAL MEANING, ORIGINAL INTENT, AND JUDICIAL REVIEW (1999); LAW AND INTERPRETATION: ESSAYS IN LEGAL PHILOSOPHY (Andrei Marmor ed., 1995).

that the legislative decision must prevail. Otherwise, courts must answer the question of severability on other grounds. I suggest below that decisions about severability and as-applied decision making should take into account the effect of severance and related practices in establishing new rules, and the relative competence of courts and legislatures to perform the rulemaking function. To explain, I begin with a brief outline of the nature and function of rules.

II. SEVERABILITY AND RULES

A. Justification of Legal Rules

A discussion of rules must begin by defining the term *rule*.[49] For my purposes, a rule is a directive issued in determinate form.[50] Laws range along a continuum from quite determinate rules to standards whose meaning cannot be grasped without an exercise of moral judgment.[51] A standard identifies a moral ideal, but leaves the practical requirements of that ideal undefined and unsettled. At the determinate end of the continuum, legal rules serve the additional function of settling controversy over the practical implications of whatever end has been selected, and guiding conduct accordingly.[52]

For example, suppose that a society has determined, through its legislative body, that distributive justice requires regulation of rental markets. The legislature might respond with a standard of conduct, "no unfair rent increases." Alternatively, it might enact an ordinance prohibiting annual rent increases in excess of increases in the Consumer Price Index ("CPI")—a highly determinate rule.

A rule, then, is based on some less determinate moral principle, but restates the principle in concrete terms.[53] The justification for choosing a determinate rule over a standard that refers more directly to the underlying moral ideal lies in the capacity of rules to prevent errors.[54] The benefits of

49. For a comprehensive analysis of the nature of and justifications for rules, *see* Frederick Schauer, PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING (1991).

50. *See id.* at 53–62 (defending the possibility of "semantic autonomy"); Frederick Schauer, *Formalism*, 97 YALE L.J. 509, 520–32 (1988) (defending the capacity of rules to constrain decision making).

51. On the distinction between rules and standards, *see generally* Cass R. Sunstein, LEGAL REASONING AND POLITICAL CONFLICT 21–34 (1996); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992); Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 HARV. L. REV. 1685 (1976); William Powers, Jr., *Structural Aspects of the Impact of Law on Moral Duty Within Utlitarianism and Social Contract Theory*, 26 UCLA L. REV. 1263 (1979); Carol M. Rose, *Crystals and Mud in Property Law*, 40 STAN. L. REV. 577 (1988).

52. *See* Alexander & Sherwin, *supra* note 48, ch. 1.

53. *See* Schauer, *supra* note 49, at 54 (rules are "instantiations" of background principles).

54. *See* Alexander & Sherwin, *supra* note 48, ch. 1 at 6–9, ch. 5 at 5–6; Joseph Raz, THE MORALITY OF FREEDOM 70–80 (1986); Schauer, *supra* note 49, at 149–55.

rules can be sorted roughly into two categories: epistemic guidance and coordination.

The epistemic benefits of rules lie in their capacity to reduce errors that occur because individual actors lack information or expertise. Armed with only a moral standard, such as distributive justice or "fair" rent, people may be unable to make correct judgments because they do not have enough economic facts. In contrast, a rulemaking body may have fact-finding capacities that give it an epistemic advantage over most actors. If, after gathering information and listening to views, the rulemaker has superior information, actors may do better by following its rule than they would do if they relied on their own judgment.[55]

A second type of error results from lack of coordination. Actors may err in applying a standard such as distributive justice because what that standard requires of them depends on the conduct of others, and they have no means of anticipating what others will do. A rule, if generally followed, makes the behavior of others predictable and puts an end to the argument and controversy that come from uncertainty.[56]

The term *coordination* is used broadly here, to include the settlement of standards of conduct that may be morally controversial but are necessary to a stable social order. Thus, in the simplest sense of *coordination,* an ordinance limiting rent increases to increases in the CPI provides coordination—even if it does not represent what is fair in every case—because it tells property owners what their legal duties are and what they can expect from officials. Moreover, at a higher level of generality, the determination that distributive justice requires rent control, or that distributive justice is a social good, also provides coordination because it settles (legally) an area of moral controversy and uncertainty.[57]

Of course, most legal rules represent (or aspire to represent) a combination of legislative expertise and coordination: There is a need to designate some course of action, and the legislature has reason to believe that one course is better than others. However, although rules can provide benefits

55. For sources recognizing the expertise function of rules, *see, e.g.*, Tom Campbell, THE LEGAL THEORY OF ETHICAL POSITIVISM 51, 58 (1996); Schauer, *supra* note 49, at 150–52, 158–59; Jules L. Coleman, *Authority and Reason,* in THE AUTONOMY OF LAW 287, 305 (R. George ed., 1996).

56. On the coordination function of rules, *see, e.g.*, Raz, *supra* note 54, at 49–50; Schauer, *supra* note 49, at 163–66; Heidi M. Hurd, *Justifiably Punishing the Justified,* 90 MICH. L. REV. 2203, 2293–3201 (1992); Gerald Postema, *Coordination and Convention at the Foundations of Law,* 11 J. LEGAL STUD. 165, 172–86; Donald H. Regan, *Authority and Value: Reflections on Raz's* Morality of Freedom, 62 S. CAL. L. REV. 995, 1006–10 (1989). On the requirement that the rule must be generally followed in order to have a coordinating effect, see Alexander & Sherwin, *supra* note 48, ch. 5 at 20–23.

57. Whether people accept the settlement is another matter. *See generally* Alexander & Sherwin, *supra* note 48, ch. 5 (discussing the difficulty of enforcing authoritative rules); Heidi M. Hurd, *Challenging Authority,* 100 YALE L.J. 1611 (1991) (rejecting the "practical authority" of legal rules). Without acceptance by those who might otherwise disagree about ends, means, or particular applications, rules have no coordinating effect. Here I assume that, in general, the rules are accepted.

in the form of coordination and expertise, they also can produce errors in several ways. Most obviously, the rule may be poorly designed, due to the rulemaker's faulty judgment or lack of expertise. A rule limiting rent increases may not in fact benefit tenants as expected, and a shift in wealth to tenants may not serve distributive justice. If so, the rule will not serve the desired moral end.

More subtly, even a good rule can, and almost certainly will, produce errors in some of its applications.[58] Rules are blunt, because they do not restate the moral principle on which they depend, but instead reduce these principles to a set of instructions about what to do in specified circumstances. As a result, some of the rules' applications will depart from the principles that support them. Even if we assume that limiting rent increases to increases in the CPI will, overall, promote distributive justice, it is not likely to promote distributive justice in every case.

Another type of rule-induced error, related to the first two, occurs when the rule provokes strategies of circumvention, with unintended consequences. Because a rule is a fixed directive that operates prospectively on a range of events that have not yet occurred, it can create incentives that alter the course of those events. Typically, there will be some actors who, for reasons of self-interest or moral dissent, do not agree with the moral premises underlying a rule. And even if we assume that all actors share the rulemaker's moral premises and want to do what is morally right, there will be those who believe, rightly or wrongly, that the rule is mistaken in its application to them. Rather than disobey and incur sanctions, these actors may alter their conduct in an effort to avoid the rule.

Problems of this kind result from a combination of bluntness and poor design. The determinate form of the rule makes circumvention possible: If actors were governed directly by the standard "Act so as to provide distributive justice," they might disobey, but they could not avoid the law. In contrast, the distributive consequences of a rent control ordinance are comparatively easy to avoid, namely, by shifting to other uses of property. If this tactic was not anticipated and if it ultimately produces a detriment in terms of distributive justice, the rule may do more harm than good.

At least in theory, there can be perfect rules. If two different courses of action are available to actors, and there are good reasons why everyone should act alike, but no good reasons to prefer one alternative over another, a rule that designates one course of action and prohibits the other will serve its moral end in every case to which it applies, and no one will seek to avoid the rule.[59] The standard example is a rule requiring that in two-way traffic,

58. *See, e.g.*, Schauer, *supra* note 49, at 31–34, 47–52 (underinclusiveness and overinclusiveness).

59. *See* Alexander & Sherwin, *supra* note 48, ch. 4 at 15–16 (equivocation rules); Regan, *supra* note 56, at 1025–26 (explaining that such enactment of a rule provides actors with a reason to believe that others will do as the rule requires, which in turn provides a reason to act accordingly).

everyone must stay to the right. But pure coordination rules are rare, and most determinate rules are subject to errors in design and errors of bluntness, and to the related problem of circumvention.

Thus, determinate rules have both benefits that arise from expertise, coordination, and settlement, and costs that arise from poor design and bluntness. It follows that the choice of a determinate rule is justified when, and only when, the rule has net compliance benefits: Judged by the underlying reasons on which the rule is based, fewer errors will occur if everyone complies with the rule than would occur if everyone judges independently what he ought to do.[60] In other words, rules serve moral ends indirectly, by providing an imperfect set of instructions that works better, overall, than a series of judgments about what to do in particular cases.[61] I do not mean to suggest that this is in fact what motivates legislators to enact determinate rules, or that courts should police statutes to ensure that they produce net compliance benefits.[62] But these are the conditions under which a determinate rule is preferable to a standard or to no rule at all.

A standard, in contrast, may be useful when the criteria for adopting a determinate rule are not met. If the legislature has settled on an objective or moral principle, but does not have sufficient information to draft a rule that will prevent more error than it causes, a standard allows individual actors, and judges as they adjudicate particular disputes, to determine the best means of carrying out the principle it expresses.[63] Over time, a series of decisions interpreting the standard may give it more determinate form.

A final general point about legal rules is that they include both primary rules of conduct and secondary rules that identify what counts as an authoritative rule of conduct and provide for the amendment of primary rules.[64] Rules that identify the legislature and govern its decision-making processes, for example, are secondary rules. On a positivist view of law, the authority of both primary and secondary rules rests ultimately on their acceptance by those who are subject to the rules.[65]

60. *See* Raz, *supra* note 54, at 70–80 (discussing the "normal justification" of rules); Schauer, *supra* note 49, at 149–55; Larry Alexander & Emily Sherwin, *The Deceptive Nature of Legal Rules*, 142 U. PA. L. REV. 1994–98; Coleman, *supra* note 55, at 304–05.

61. *See* Larry Alexander, *Pursuing the Good—Indirectly*, 95 ETHICS 315 (1985).

62. Public choice theory, for example, raises serious doubts about orderly pursuit of public good through legislation. *See generally*, Daniel A. Farber & Philip P. Frickey, *The Jurisprudence of Public Choice*, 65 TEX. L. REV. 873 (1987); *Symposium on the Theory of Public Choice*, 74 U. VA. L. REV. 167 (1988).

63. *See* H.L.A. Hart, THE CONCEPT OF LAW 121–23 (1961) (discussing the use of standards).

64. *See id.* at 77–96 (discussing secondary rules in a mature legal system).

65. *See* Joseph Raz, THE AUTHORITY OF LAW 42–43 (1979) (efficaciousness). At a minimum, officials must accept the rules and subjects must accept the basic secondary rule that gives officials their authority. *See* Hart, *supra* note 63, at 94 (on the "rule of recognition").

B. The Role of Rules in Judicial Review

1. A Rule-Oriented Analysis of Severability

Part I of this article describes a collection of principles that work to preserve parts of legislation in the course of adverse judicial review. Courts may uphold a statute as applied despite doubts about other applications, or hold the statute unconstitutional as applied and leave it otherwise in force; they may hold that invalid applications or the statute are severable from valid applications; or they may provide a saving construction that they would not endorse in the absence of constitutional problems. In each case, the product of the court's decision is a new law.

The first conclusion to be drawn from this observation is that the new law that will result from severance or similar judicial action should be judged *as a rule*. It is not enough that severance will enable the court to reach a desirable result; nor should the court assume that partial enforcement of legislative policy is better than no enforcement at all. Because severance or a saving interpretation replaces one rule with another, the use of either should depend on whether the new rule that results is consistent with the original rule and justified in its own right.

There are two parts to this test. First, the principle or goal served by the original statute should remain constant, in deference to the legislature's primary role in settling public policy. Second, the new rule should be justified in the sense that, judged by the original principle or goal, it will produce net compliance benefits. In other words, the rule must prevent more errors of judgment through coordination and expertise than it causes through poor design or bluntness.

Notice that "legislative supremacy," in the sense of hierarchical authority, plays no part in the critical question whether the rule is a justified rule. Nor should it, because either of the available courses of action—invalidating the entire statute or upholding a part of the statute—can be characterized as an act of judicial deference to the legislature (or as an act of judicial aggression). One can say that courts respect legislative prerogative when they preserve what they can of partly invalid statutes; one can also say that courts respect legislative prerogative when they decline to substitute new, narrower rules in place of the rules originally enacted. In fact, the court is acting aggressively in both cases, either by eradicating legislation or by replacing it with a new rule.

Because the court cannot defer, it must face the question of rulemaking and ask how the remaining portions or applications of the statute will function as a rule. This inquiry can be characterized as an inquiry into hypothetical legislative intent: Given the failure of part of its statute, a legislature presumably would approve of a rule that produced net benefits and disapprove of a rule that did not. It seems more forthright, however, to think of what the court is doing as an independent analysis carried out in the absence of any discernible legislative intent.

For a simple example of how a rule-oriented analysis might work, suppose that a rent control ordinance limits rent increases to increases in the CPI, prevents termination of leases as long as rent is paid, and flatly prohibits conversion of rental property to other uses. This ordinance is challenged by a landlord who has no plans for conversion. The court concludes that the basic rent control provision is valid. It also concludes—improbably[66]—that the unqualified ban on conversion amounts to an unconstitutional taking of property.

In these circumstances, the court can find the statute severable and enforce it as applied to the challenger, or find it nonseverable and therefore unenforceable in any of its applications. Severance will prevent the landlord who initiated the case from raising rent—a satisfying result if one accepts the legislative premise that rent control is just. On a rule-oriented view, however, the question is not whether the result is in accord with legislative intent, but whether the new rule of rent control without limits on conversion is a sound rule. This is less obvious: Under some economic conditions, the new rule might encourage rapid conversions from rental property to condominium or commercial use, an undesirable result if one accepts the legislative premise. Thus, the problem looks somewhat different when treated as a problem of rules than as one of results.

Another example is suggested by the aftermath of *Metromedia, Inc. v. San Diego*.[67] San Diego had enacted an ordinance that banned most billboards and required existing billboards to be retired over time. The United States Supreme Court held that the ordinance was valid insofar as it banned commercial billboards but invalid with respect to noncommercial billboards.[68] It then remanded the case to the California Supreme Court for consideration of severability or saving interpretation.

66. The Supreme Court has made clear that rent control is not, in itself, an unconstitutional taking of private property. See Pennell v. San Jose, 485 U.S. 1 (1988) (finding a taking claim premature when tenant hardship was only one among seven factors to be considered in approving rent increases). In Yee v. Escondido, 503 U.S. 519 (1992), the Court upheld a set of regulations on mobile-home pad rentals that controlled rents and also limited landlords' choice of tenants, limited the grounds for terminating leases, and prevented landlords from charging fees or removing mobile homes in case of sale. The Court suggested, however, that if the law had prohibited termination of leases and conversion of property to other leases, either explicitly or as applied, this might amount to a "physical" taking. *Id.* at 528. It also suggested that strict limits on landlords' ability to choose among tenants could result in a regulatory taking, but then strained to conclude that the issue of regulatory taking was not properly before the Court. *Id.* at 530–31, 535–38.

67. 453 U.S. 490 (1981), *on remand to the Supreme Court of California* 32 Cal. 3d 180 (1982).

68. *Metromedia, Inc.*, 453 U.S. at 500–515. The California Supreme Court had held that the ordinance was valid on its face but invalid as applied to noncommercial billboards when there were no good alternatives open to particular speakers. Metromedia, Inc. v. San Diego, 26 Cal. 3d 848, 869 n.14 (1980). The United States Supreme Court held the ordinance invalid with respect to all noncommercial speech at least when the ordinance as written preferred commercial to noncommercial speech and attempted to regulate the content of noncommercial billboards by allowing some types of noncommercial displays and prohibiting others. *Metromedia, Inc.*, 453 U.S. at 512–515.

The California court determined that the ordinance was plainly intended to cover noncommercial billboards, and that, despite a general severability clause, its applications to commercial billboards could not be severed from its applications to noncommercial billboards.[69] The reasoning of the California court tracks, at least loosely, the analysis I have suggested. The court might simply have concluded that any billboard removed from the road is a partial realization of the legislative goal, therefore the ordinance should be severed. Instead it considered how the new rule (banning commercial billboards only) would operate as a rule. The new rule might encourage a shift in billboard use from commercial to noncommercial messages, and would require municipal enforcers to police the content of rapidly changing billboards.[70] It also might be inferior to a rule regulating the location and appearance of all billboards.[71] Therefore, there was no assurance that the rule resulting from severance would be a sound rule, and a good chance that the legislature might have done something better.

I do not mean to suggest that the rule-oriented analysis I have described represents a dramatic new approach to severability and related issues. At least when the issue is characterized as one of severability, courts may engage in roughly the analysis I have proposed when they consider whether the legislature "would have enacted" the remaining provisions of the statute. This appeared to be the case in *Metromedia*. But an explicit recognition of the effect of the decision as creating a new rule, coupled with an understanding of the circumstances under which rules are justified, can help to avoid the facile assumption that partial enforcement of a statute is better than no enforcement. An explicitly rule-oriented analysis may also suggest some useful distinctions among types of statutes; I will attempt to sketch some of these below. Finally, a rule-oriented analysis casts doubt on current doctrine pertaining to as-applied adjudication and saving interpretation, which leans heavily in favor of narrowing rather than invalidating legislation, and thus discourages critical consideration of the resulting rule.

Summarizing to this point: A rule-oriented approach to partially invalid statutes suggests that courts should not replace the original legislation with a new rule unless they are confident that the new rule will produce a favorable balance of error, in light of the moral principle or goal underlying the original legislation. Errors prevented through expertise and coordination must exceed errors resulting from the bluntness of the rule. If these requirements are not met, there is no reason to assume that the new rule is preferable to no rule at all.

It does not follow, however, that the court itself should examine the justification of the new rule in the manner just described. Rather, unless the net benefits of the rule are obvious, the court should first consider which body—itself or the legislature—is best positioned to issue a rule. If the court

69. Metromedia, Inc. v. San Diego, 32 Cal. 3d 180, 186–90 (1982).
70. *Id.* at 190.
71. *Id.* at 191.

refuses to sever or otherwise save a partly invalid statute, the effect of its decision is to assign the rulemaking task to the legislature. Therefore, in a case of even moderate difficulty, comparative institutional competence should be a threshold question preceding direct analysis of the justification of the rule.

Once again, the standard reference to legislative supremacy is not helpful. A decision to sever a statute may honor legislative supremacy by preserving as much legislation as possible; then again, a decision not to sever honors legislative supremacy by assuming that the legislature is the superior rulemaker. What is needed, therefore, is a direct comparison of legislative and judicial rulemaking, given that the original rule cannot be left in place.

Although legislatures are surely subject to distorting influence in their choice of ends, they ordinarily are superior to courts as *rulemaking* bodies. Rulemaking involves choosing a set of determinate criteria for action that, if generally accepted, will serve to promote a given end. To do this effectively, the rulemaker must assess how the rule will operate over a range of future cases. Because a court must judge rules from the vantage point of a single dispute, without the benefit of legislative fact-finding tools and without having heard the views of many affected parties, its judgment about the overall effects of a rule may be inaccurate. Cognitive science suggests that human minds are disposed to miscalculate probabilities because they focus on salient facts at the expense of equally relevant background conditions.[72] The process of adjudication, with its focus on the positions of particular individuals, invites this kind of bias. Moreover, at least if one accepts democratic premises, a legislature is better able to resolve moral controversy through compromise. Courts are neither authorized by political processes, nor equipped by access to public opinion, to settle contested questions.

Of course, courts do make rules, but they normally proceed in the incremental manner associated with the common law.[73] Over the course of a series of individual decisions, the courts begin to pick out rules that connect the cases and appear to be justified as rules for future cases as well. This process, which has similarities to the process of reflective equilibrium, works to counteract adjudicatory bias.[74] In contrast, when a court amends a statute through severance or a similar device, there is no opportunity for incremental development of a new rule: The court must act at once. For this purpose, courts have serious shortcomings.

72. *See, e.g.*, Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 163, 163–65, 174–78 (Daniel Kahneman, Paul Slovic, & Amos Tversky eds., 1982).

73. On the nature and virtues of analogical reasoning in law, see, for example, Anthony Kronman, THE LOST LAWYER 170–85 (1993); Cass R. Sunstein, LEGAL REASONING & POLITICAL CONFLICT 62–100 (1996); Scott Brewer, *Exemplary Reasoning: Semantics, Pragmatics, and the Rational Force of Legal Argument By Analogy*, 109 HARV. L. REV. 923 (1996); Emily Sherwin, *A Defense of Analogical Reasoning in Law*, 66 U. CHI. L. REV. 1179, (1999). For strong criticism of analogical reasoning, see Larry Alexander, *Bad Beginnings*, 145 U. PA. L. REV. 57 (1996).

74. *See* Sherwin, *supra* note 73. On reflective equilibrium, *see* John Rawls, A THEORY OF JUSTICE 46–53 (1971); Brewer, *supra* note 73, at 938–39, 963, 1023.

All else being equal, the shortcomings of courts as rulemakers would indicate that courts should not assume rulemaking responsibility through severance and related practices, but instead should return the rulemaking task to the legislature. In the context of adverse judicial review, however, legislative action has costs as well in the form of delay and disruption. The benefits (if any) of a reduced form of the statute are lost, at least temporarily. Further, complete invalidation of the statute usually brings about a greater change in law than partial invalidation. Change creates uncertainty, and uncertainty generates errors and costs as actors try to anticipate what will happen next.

Thus, as an initial step in determining whether severance or similar judicial action is the right way to proceed, courts should compare their own shortcomings as rulemakers with the potential costs of postponed legislation. There is no metric for this comparison, but it should at least be possible to judge whether disruption and delay are likely to be extensive, and to identify some contexts in which courts are more or less competent as rulemakers. For this purpose, much will depend on the characteristics of the statute at issue.

When the original, partly invalid statute is a law of general application with roots in the common law, there is a strong case for severance or as-applied adjudication. Consider, for example, a trespass statute found to be unconstitutional when applied to Jehovah's Witnesses distributing leaflets in a company town.[75] Having found one application of the statute unconstitutional, the court has a choice whether to invalidate the statute in its entirety or only as applied.[76] The effect of an as-applied decision would be to replace a very general but determinate rule ("Do not enter another's property without consent") with a slightly narrower determinate rule ("Do not enter another's property without consent, except to distribute literature in a company town").

In this case, complete invalidation of the statute would cause considerable disruption by temporarily eliminating a basic form of protection for property rights in land. Because the trespass statute is largely intact, it is also fair to assume that if the original statute was justified, so too is the revision. Moreover, the court's qualifications as a rulemaker are at their strongest because the statute itself incorporates a body of property law developed by courts in the manner of the common law.

Judicial revision of statutes through severance, as-applied adjudication, or narrowing interpretation may also be warranted if the rule in question is based mainly on coordination rather than expertise. In other words,

75. Marsh v. Alabama, 326, U.S. 501. I am assuming that the threat this statute poses to other First Amendment rights is insubstantial, so that the question of severability has no constitutional dimension. On this point see Alexander, who argues for a balance between the deterrent effect the statute may have on expression and the legislative interest in regulating trespass by means of a broad prohibition. Alexander, *supra* note 7, at 552–54 (1981).

76. A saving construction is implausible in this type of case.

some rule is necessary, but not much turns on which rule is selected. Suppose that the legislature enacts a traffic rule that imposes penalties for "driving on the wrong side of the road," and the court finds the law to be unconstitutionally vague.[77] Despite the traditional hostility to judicial line-drawing and "rewriting" of statutes, it seems well within the court's competence to fix this statute by choosing "left" as the prohibited side. For coordination, what is needed is not expertise or unbiased fact-finding about a field of potential applications of the rule, but only a choice by someone with authority. If there is a reason for judicial reluctance to make this kind of revision, it is the risk of mistaking a pure coordination problem for one that involves a choice among unequal and controversial alternatives.[78]

In contrast, when the new rule depends for its justification on the rule-maker's expertise or on settlement of contested moral questions, and the original statute was not a comprehensive enactment of common law, a court should be wary of exceeding its rulemaking competence. For example, a decision to sever an invalid anticonversion provision from an otherwise valid rent control ordinance requires a determination that rent control without conversion limits will not be self-defeating. Similarly, a decision to read a rule "Do not drive too fast" to mean "Do not drive more than 55 miles per hour" requires a determination that a desirable speed for a safe and efficient traffic flow is 55. A decision to uphold a child care statute, but to sever an unconstitutional provision allowing parents to apply federal funds to religious day care, requires a determination that provision of some funds for day care justifies favoring secular over religious providers. Given the inferiority of courts as rulemakers—their retrospective adjudicatory bias, their limited fact-finding processes, and their lack of political authority to settle moral controversy—the current assumption that courts ordinarily should make these determinations seems out of place. In particular, the strong presumption in favor of as-applied adjudication and saving interpretation is unjustified.

The same considerations suggest that it is not a good idea for legislatures to include blanket severability clauses in statutes. General severability clauses are likely to have been enacted with no specific constitutional challenge in mind, and so cannot be viewed as deliberate exercises in backup rulemaking by legislatures. The best argument for judicial enforcement of severability clauses is that the court must honor all explicit provisions in order to maintain a general rule of judicial respect for statutory

77. Criminal statutes deny due process of law when they are too vague and uncertain to provide fair notice of what is required. *See, e.g.*, Larzetta v. New Jersey, 306 U.S. 451, 458 (1939) ("A statute which either forbids or requires the doing of an act in terms so vague that men of common intelligence must necessarily guess at its meaning and differ as to its application, violates the first essential of due process of law.").

78. In effect, it may be wise to adopt a *rule* against judicial rewriting of statutes because, although the rule is inapt in pure coordination cases, it prevents mistakes of judgment about when expertise is needed, and thus prevents more errors than it creates.

text.[79] In a great many cases that have nothing to do with severability, the rule of respect for the text is sound because it recognizes the comparative shortcomings of courts as rulemakers; and perhaps a practice of ignoring the text in severability cases is subversive of the general rule. But from the legislative point of view, severability clauses seem an unwise delegation of rulemaking power unless they refer to specific provisions and are based on a conscious consideration of the possibility of severance.[80]

Another question that may bear on the rulemaking competence of courts and legislatures is whether the rule under consideration is a primary rule of conduct or a secondary rule—that is, a rule governing the issuance and authority of conduct rules. When severability or related questions arise in connection with a secondary rule, the problem of justifying a new rule is more complex, though similar in substance. A revised secondary rule is justified if it will result in conduct rules that yield net benefits.

As an example, consider the problem of legislative vetoes.[81] Before 1983, a large number of federal statutes delegated rulemaking power to agencies, but provided for a veto of executive rules by one house of Congress. The Supreme Court finally held these vetoes unconstitutional, creating a widespread problem of severability.[82] Perhaps sensing that much disruption would follow if it invalidated complex legislation on matters such as immigration and airline deregulation, the Court decided several cases in favor of severability, finding that Congress "would have enacted" the statutes without vetoes.[83] Severance of an invalid legislative veto, however, is a risky undertaking by the court. To justify the resulting new secondary rule, which decreases legislative oversight of agency rulemaking, it must be the case that agency rules—subject to repeal only through a full legislative process—will yield net benefits to society. A judgment of this kind is highly demanding of information and expertise, and the legislative veto provision suggests that the legislature itself had doubts about potential agency rules. Perhaps the disruption that would follow from invalidating the statutes in question was enough to overcome the risks of the judge-made secondary rule. But if that is the case, it seems better to face the comparison directly than to ask what Congress would have intended about severability.[84]

79. *Cf.* Nagle, *supra* note 2, at 234–35 (advocating a "plain meaning" rule for severability clauses, consistent with judicial treatment of other statutory provisions).

80. *See* Stern, *supra* note 2, at 122–23 (suggesting that severability clauses are not well-thought-out by legislatures). For an instance in which Congress appears to have consciously considered the effects of severability, see Nagle, *supra* note 2, at 241.

81. Severability of legislative veto provisions is discussed in Nagle, *supra* note 2, at 204–05 & n.6 (collecting other sources); Laurence H. Tribe, *The Legislative Veto Decision: A Law By Any Other Name?*, 21 HARV. J. ON LEGIS. 1, 21–27 (1984).

82. INS v. Chadha, 462 U.S. 919, 944–59 (1983).

83. Alaska Airlines v. Brock, 480 U.S. 678, 684–85 (1987); *Chadha*, 462 U.S. at 931–35, 959 (relying on severability clause).

84. John Nagle reaches the contrary conclusion that the speculative nature of hypothetical Congressional intent supports a focus on actual intent with respect to severability, coupled with a plain statement rule in favor of severability. Nagle, *supra* note 2, at 230, 232–58.

2. Formal Considerations

In formal terms, most of the examples discussed in the previous section involved the replacement of a determinate statute with a new rule that is still in determinate form. Severance of an invalid anticonversion provision from an otherwise valid rent control ordinance, for example, results in an equally determinate new rule. The formal consequences of severance, as-applied adjudication, and narrowing interpretation, however, do not always follow this pattern.

One variation occurs when a court revises a determinate rule by introducing a narrowing standard. For example, the court might hold that the rent control ordinance is valid except as applied to landlords suffering economic hardship. The result is considerably further toward the standard-end of the formal continuum than is the original law, because *hardship* is a controversial term entailing questions about distributive justice.

In this situation, there is a special reason for caution, beyond the question of judicial competence to make rules. The court is changing not only the substance of the rule, but its form. A determinate rule represents a decision by the legislature to regulate by rule. As explained earlier, a determinate rule may produce error because it is blunt, but it also can prevent errors of judgment by actors and courts, or other bodies charged with administering the rule. If determinacy is important to the ends of the legislation, the court should either try to find a determinate alternative (subject to considerations of competence), or invalidate the statute in its entirety and return the rulemaking task to the legislature.

In other cases, the original legislation may take the form of a standard, which the court replaces with something closer to a determinate rule. Here, the analysis changes. Suppose, for example, that a rent control ordinance imposes criminal penalties for "unreasonable rent increases." The statute is challenged by a landlord who has quadrupled rent in an effort to drive out existing tenants, on the ground that it is unconstitutionally vague. The ordinance undoubtably is vague, but there is a qualification to the constitutional requirement of clarity in penal statutes, to the effect that one whose conduct is clearly covered cannot raise the claim of vagueness.[85] Therefore, the court might agree that the ordinance is too vague but uphold it as applied to this landlord, on the ground that he must have known it covered sudden and very large increases designed to drive out tenants. When it does this, the court moves the statute in the direction of determinacy. We now

85. *See, e.g.*, Village of Hoffman Estates v. The Flipside, Hoffman Estates, Inc., 455 U.S. 489, 494–95 (1975) ("A plaintiff who engages in some conduct that is clearly proscribed cannot complain of the law as applied to others."); United States v. National Dairy Products Corp., 372 U.S. 29, 31 (1963) (upholding a provision of the Robinson-Patman Act that forbade the selling at "unreasonably low prices" as applied to a dairy that sold milk below cost) ("In determining the sufficiency of notice, a statute must of necessity be examined in light of the conduct with which the defendant is charged."). When vague statutes affect the exercise of First Amendment rights, the court has reached the opposite result and allowed facial challenges. *See, e.g.*, Coates v. Cincinnati, 402 U.S. 611 (1971) (conduct "annoying to persons passing by").

know one type of case that is definitely covered. And it is possible that over time, the statute will gain a body of determinate content in the course of similar decisions.

This seems a particularly strong case for judicial rulemaking. Recall that a standard is preferable to a rule when the rulemaker has in mind an end or value to be served, but does not yet have the information or expertise necessary to reduce that end to a determinate rule that will prevent more error than it causes.[86] Thus, when a legislature chooses to enact a law in the form of a standard, it may lack the information needed for a rule. In these circumstances, the best way to acquire information and expertise about specific cases that fall within the scope of the standard may be to proceed in the incremental manner of the common law, deciding concrete cases as they arise and developing more concrete rules over time. In effect, the initial standard is an invitation to the court to move in this way toward a rule.

The court's competence to revise a statutory standard, however, depends on this incremental mode of decision making. The court is certainly no better equipped than the legislature, and probably less well equipped, to issue a rule at the outset. Thus a process of gradual narrowing through as-applied decisions that identify permissible and impermissible applications is preferable to a comprehensive saving construction. The court should avoid holding immediately that "unreasonable rent increases" means "rent increases in excess of the landlord's costs."

3. The Possibility of a Rule

The preceding sections suggest a variety of considerations that might enter into a court's analysis of whether to sever or otherwise alter a particular statute in the course of adverse judicial review. One alternative to an analysis of this kind would be to establish a flat rule one way or the other: "Always sever," or "Never sever," or perhaps "Sever unless the statute targets expression." Although the Supreme Court has preferred case-by-case decision making to a rule on the question of severability, some of its statements on as-applied adjudication and narrowing interpretation come close to establishing a rule in favor of these practices.[87]

A rule for or against severability, as-applied adjudication, or narrowing interpretation must meet the general standard for justification of rules. That is, the rule is justified if it will prevent more errors of judgment by courts in particular cases than it will cause by preempting correct decisions in particular cases. Although a rule would certainly simplify decision making, the

86. There are reasons apart from lack of expertise why a legislature might choose a relatively open-ended standard. For example, a standard deters actors from operating close to the line between acceptable and unacceptable behavior. Over time, however, a highly indeterminate standard cannot provide the benefits of settlement and guidance, nor produce consistency in adjudication, because it leaves too much room for judgment on morally controversial questions. Therefore, courts will naturally tend to move standards in the direction of determinacy.

87. *See, e.g.*, United States v. Salerno, 481 U.S. 739, 745 (1987) (as-applied adjudication); De Bartolo Corp. v. Florida Gulf Coast Building & Construction Trades Counsel, 485 U.S. 568, 575 (1988) (saving construction).

relative rulemaking competence of courts and legislatures varies sufficiently with different statutes that individualized decision-making appears preferable to a flat rule. If a rule were adopted, however, concerns about the competence of courts as rulemakers suggest that a rule against severance and similar narrowing practices would be wiser than a rule in favor.

CONCLUSION

Invalidating a statute is an imperious act by a court, with serious consequences. It is not surprising, therefore, that courts have tried to preserve statutes that are only partly invalid by severing invalid provisions or applications, limiting their decisions to particular applications of the statute raised by a case, or providing saving interpretations they would not otherwise adopt. A rule-oriented view of law, however, suggests that these practices are not as deferential as they appear, and may not always be desirable.

When a court avoids full invalidation of a statute by severance, as-applied adjudication, or saving interpretation, the effect of its decision is to create a new law, usually in the form of a new determinate rule. Once this is understood, several conclusions follow. First, these three practices are similar in effect and should be judged by similar criteria. Second, courts should not engage in these practices uncritically on grounds of "legislative supremacy." Legislative supremacy is unhelpful because the alternative to judicial invalidation of statutes is judicial rulemaking. Third, whether the new rule that results from severance and related practices is a justified rule depends on whether it yields net benefits, when judged by the principle or goal the original statute was designed to promote. Unless the net benefits of the rule are evident, the court should begin by considering comparative institutional competence to make this judgment, and if the legislature appears to be better positioned for this purpose, should decline the rulemaking responsibility. In general, courts are poorly situated to issue comprehensive, prospective rules; but their shortcomings as rulemakers will sometimes be overcome by the delay and disruption that will follow if the rulemaking function is returned to the legislature.

For these reasons, the current presumptions in favor of severability, as-applied adjudication, and saving interpretation of statutes should be dropped. This would mean that more legislation is completely undone by judicial review. But it seems preferable to practices that obscure the full consequences of judicial review by purporting to be modest when, in fact, they result in new, judicially created rules. It also aligns the practice of judicial review with the view that the Constitution is not a code of protected conduct but a set of constraints on legal rules—a view I find persuasive.[88]

88. *See* Adler, *supra* note 5, at 153–58 (arguing that a "derivative account" of constitutional rights, in which courts address the constitutionality of rules, indicates that courts must engage in facial review of statutes).